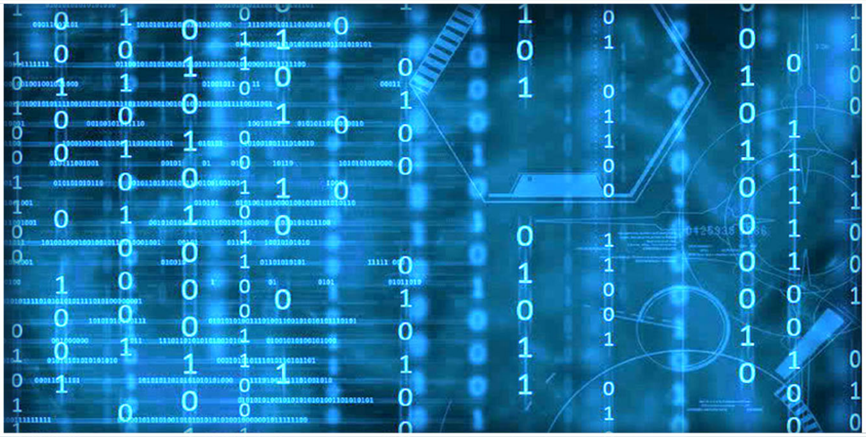
IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Volume 13 Issue 7

July 2022



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 13 Issue 7 July 2022
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

CONTENTS

Research Progress and Trend of the Machine Learning based on Fusion

Chen Xiao Yu[1], Song Ying[3]
Computer School, Beijing Information Science & Technology University, Beijing, China

Zhang Xiao Min[2], Gao Feng[4]
Academy of Agricultural Planning and Engineering
Ministry of Agriculture and Rural Affairs, Beijing, China

***Abstract*—Machine learning is widely used in the data processing including data classification, data regression, data mining and so on, and based on a single type of machine learning technology, it is often difficult to meet the requirements of data processing; in recent years, the machine learning based on fusion has become an important approach to improve data processing effect, and at the same time, corresponding summary study is relatively limited. In this study, we summarize and compare different types of fusion machine learning such as ensemble learning, federated learning and transfer learning from the perspectives of classification, principle and characteristics, and try to explore the research development trend, in order to provide effective reference for subsequent related research and application; furthermore, as an application of fusion machine learning，we also conduct a study on the modeling optimization for car service complaint text classification.**

Keywords—Machine learning; fusion; ensemble learning; federated learning; transfer learning

I. INTRODUCTION

In recent years, machine learning research has developed rapidly and achieved widespread attention; with the expanding and deepening of the development, it is found that traditional machine learning methods often fail to meet the needs of data modeling for certain scenarios; therefore, fusion-based machine learning has become a research hotspot, and relevant research has involved the fields of agriculture, geology, environment, machinery, communication, medicine and so on. This study summarizes and compares the fusion-based machine learning methods from multiple perspectives, it is expected that this study could provide support for effectively obtaining a general understanding for the research progress of the machine learning based on fusion, and explore the research development trend based on summary and analysis.

II. MACHINE LEARNING BASED ON FUSION

Fig. 1 summarizes the main fusion-based machine learning methods from the perspectives of classification, principles, and characteristics.

Fusion-based machine learning involves technology fusion and data fusion.

The technology fusion includes horizontal fusion and vertical fusion. The horizontal fusion mainly refers to ensemble learning, the ensemble strategies include bagging, blending, stacking and so on. The vertical fusion mainly refers to different types of neural networks stacking to form multiple neural networks, and also includes other multi-layer information processing methods based on the stacking of different types of basic technologies.

The Data fusion mainly refers to federated learning which is based on distributed model training; and the method of transfer learning shares model training results through parameters between the models trained based on different parts of data, which could be understood as another kind of machine learning based on data fusion.

A. Ensemble Learning

Ensemble learning commonly uses multiple algorithms to train individual models independently of each other, and then combines the training results through certain strategy to form a comprehensive model based on model horizontal combination to improve the modeling effect.

Different types of machine learning technologies commonly have different characteristics, applicability and limitation, and ensemble learning could effectively combine the advantages of multiple machine learning technologies to optimize the effects of machine learning modeling. In recent years, ensemble learning has received extensive attention, related research had involved agriculture, geology and so on [1-2].

The basic technologies used in the individual model training mainly include the classic technologies such as random forest, support vector machine, ridge regression, and the technologies based on neural network such as bidirectional long short term memory neural network, error back propagation neural network and so on; the ensemble strategies mainly include bagging, blending and stacking.

The bagging ensemble strategy is mainly based on the idea of voting, for classification issue, the calculation results from the individual models could be regarded as votes, and the prediction category with the most votes could be took as the classification prediction output of the overall ensemble model, for regression issue, the weighted average of the outputs from multiple individual models could be computed as the output of the overall ensemble model.

Funding Project: Promote the Classification Development of College-Construction of Professional Degree Sites of Electronic Information (Computer Technology) (5112211038).

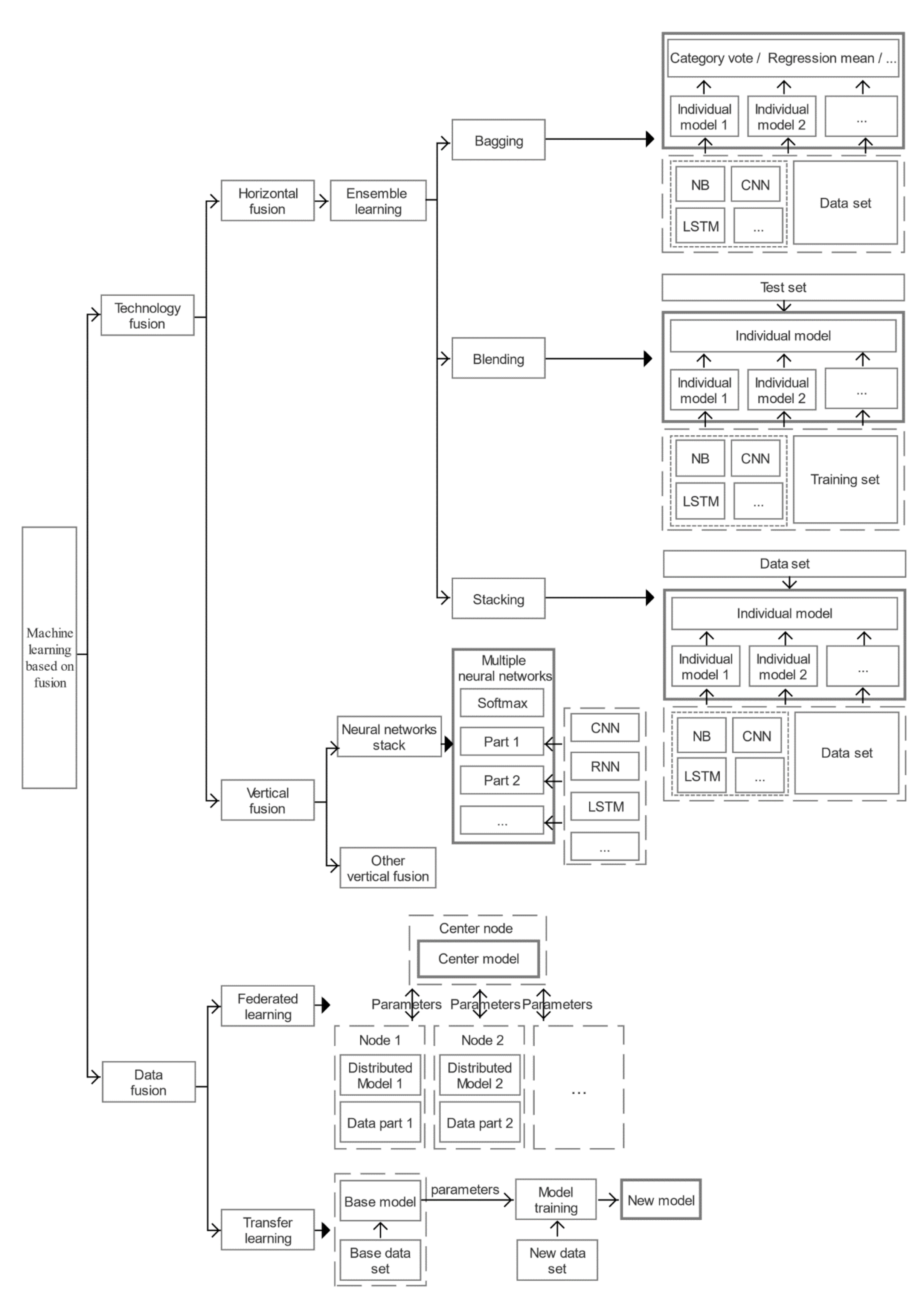


Fig. 1. Machine Learning based on Fusion.

The blending ensemble model usually consists of two layers, the first layer is commonly a plurality of individual models trained independently, the second layer is another individual model, which operate prediction based on the outputs of the first layer, and the predicted result from the second layer would be used as the final output of the ensemble model; to build a blending ensemble model, the dataset is commonly divided into a test set and a training set firstly, and then the model training of first layer is performed based on the training set, and the model training of second layer is performed based on the test set.

The stacking ensemble model commonly includes two layers too, the first layer consists of multiple individual models; and the second layer is another individual model, which conducts prediction based on the outputs of the multiple individual models of the first layer; the stacking ensemble strategy is similar to the blending ensemble strategy, the main difference lies in the data segmentation in the model training, in stacking model training, the data used for the first-layer model training and the second-layer model training is consistent, and in the training of blending model, the training data of the second-layer does not intersect with the training data of the first-layer.

Table I generally summarizes and compares the research examples for the ensemble learning from the aspects of ensemble strategy, basic technology, etc.

B. Transfer Learning

Transfer learning commonly initializes new model training based on pre-trained model parameters, and through little applicability adjustment, achieves quickly and effective data modeling [3-6]. The transfer learning methods used in related research are mainly based on neural network, which commonly redesigns the fully connected layer and freeze the other layers of the pre-trained model, or freeze part layers and adjust the other layers.

Transfer learning provides an effective alternative approach mainly for two kinds of scenarios, one is the data modeling based on a limited amount of data, by transfer learning, effective modeling could commonly be achieved through fine-tuning based on a pre-trained model trained based on enough data; the second is the scenario where the machine learning model training takes too much time, through transfer learning, the pre-trained model could be used as basis to quickly realize the modeling for a new dataset, which could effectively reduce the training time and reduce the timeliness constraint of the model use.

Transfer learning is commonly based on convolutional neural networks, and related methods mainly include four categories, including the methods based on AlexNet, the methods based on VGGNet, the methods based on ResNet and the methods based on DenseNet. Table II generally summarizes and compares the research examples for different kinds of transfer learning from the aspects of research method, benchmark method and so on.

TABLE I. FUSION STRATEGIES OF ENSEMBLE LEARNING AND THE RESEARCH EXAMPLES

No.	Integration strategy	Research method	Research object	References
1	Bagging	Individual learning: KNN (k-nearest neighbor), BP (error back propagation neural network), GBDT (gradient boosting decision tree), RF (random forest); using improved weighted voting strategy	Electricity theft detection	[7]
		Individual learning: DT (decision tree)	Personal credit evaluation	[8]
		Individual learning: DT	Character recognition	[9]
		Obtaining different datasets based on the bootstrap sampling method and training multiple BP models separately	Water bloom prediction	[10]
2	Blending	Individual learning: GBDT, linear-SVM (linear support vector machine), RBF-SVM (radial basis function-support vector machine), Fusing: linear-SVM	Infrared spectroscopy data analysis	[11]
3	Stacking	Individual learning: LR (logistic regression), KNN, RF, GBDT Fusing: GBDT	Strains classification of Anoectochilus roxburghii	[12]
		Individual learning: RF, XGBoost, LightGBM Fusing: LR (linear regression)	Gaseous nitrous prediction (in air)	[13]
		Individual learning: RF, GBDT, SVM (support vector machine), RR (ridge regression) Fusing: RR	Estimation of summer corn fractional vegetation coverage (based on drone multispectral image)	[14]
		Basic method: KNN, SVM, RF, GBDT	Classification of rice phenomics entitie	[15]
		Basic method: KNN, RF, adaptive boosting	Estimation of nitrogen contents in citrus leaves	[16]
		Basic method: RF, LDA (latent dirichlet allocation), LIBSVM	Classification of black Goji berry	[17]

TABLE II. CATEGORIES OF TRANSFER LEARNING AND THE RESEARCH EXAMPLES

No.	Category	Research method	Contrast method	Research object	References
1	AlexNet	AlexNet	SVM (support vector machine, BP (error back propagation neural network), VGG-19, GoogLeNet Inception v2	Image recognition of cotton leaf diseases and pests	[18]
		AlexNet	CNN (convolutional neural network)	Flotation performance recognition	[19]
2	ResNet	ResNet-101	ResNet-50	Intelligent lithology identification	[20]
		TL-SE-ResNeXt-101 (based on improved deep residual network SE-ResNeXt-101)	ResNet-50, VGG-16, DenseNet-121, GoogLeNet	Crop disease classification	[21]
		CDCNNv2 (based on residual network ResNet-50)	ResNet 50, VGG16, VGG19, DenseNet 121, Xception	Crop diseases detection	[22]
		MPDE-VMD+DTL (multiple population differential evolution-variational mode decomposition, deep transfer learning)	BP, ResNet, migration component analysis	Mechanical fault diagnosis	[23]
3	VGGNet	VGG-16	AlexNet, ResNet 50, Inception v3	Grape leaf disease detection	[24]
		Grape-VGG-16 (a kind of grape leaf disease identification model)	New learning	Grape leaf disease recognition	[25]
4	DenseNet	DL–T (based on DenseNet and LSTM)	RNN–T	Speech recognition	[26]
		DenseNet-GCForest (based on DenseNet and deep forest)	CNN	Wafer map defect recognition	[27]

C. Federated Learning

Federated learning commonly does not centralize training data, but is based on distributed model training, independent partial model training is performed on multiple terminals based on different parts of the data, and the partial training result parameters would be centralized, based on which the central model parameters could be updated through certain strategy, so as to achieve the purpose of integrating multi-source data to improve the machine training effect.

The characteristics and advantages of federated learning mainly include three aspects, firstly, it is not necessary for federated learning to centralize raw data from different sources or terminals, which could effectively protect the data security and user privacy; secondly, federated learning could be combined with the idea of edge computing, and a large amount of data processing workload would be distributed to terminals, which could effectively relieve the computational pressure of the central node and improve the performance of the overall machine learning system; finally, under the condition of multi-node computing, since the data transmission between the terminal nodes and the central node does not involve the transmission of original data but only a small amount of data such as training parameters, the data transmission pressure could be significantly reduced and the performance of the machine learning system could be further optimized.

With the rapid development and wide application of information technology, as the disadvantage side of the double-edged sword of the technology, data security and privacy security risk are gradually being highlighted; at the same time, the application of the internet of things is becoming more and more extensive, and there are always more and more big data processing scenarios; therefor, federated learning has become a hot topic. Related researches have involved the fields of communication, medicine, multimedia and so on. In the field of communications, Wang Jia Rui et al. (2021) proposed a clustered wireless federated learning algorithm for the scenario of high-speed internet of vehicle, the test based on handwriting recognition model show that under the condition in which the channel state is poor and the user transmit power is greatly limited, the convergence value of the loss function could be effectively reduced based on this method, compared with traditional centralized algorithm [28]. Jing Xing Hong et al. (2021) proposed a LTE-V2X (long term evolution-vehicle to everything) channel estimation algorithm based on federated learning, which estimates the time-varying channel based on CNN-LSTM-DNN (convolutional neural network-long short term memory-deep neural network), and allocates the required computation to vehicle users, research result show that the method could effectively track the time-varying channel in the high-speed mobile scene of vehicle user, which only lose a small amount of performance compared with the centralized learning algorithm [29]. In medical field, Wang Sheng Sheng et al. (2021) proposed a kind of federated learning method based on improved RetinaNet and attention mechanism for the privacy protection requirements in the model training for medical image target detection, and research result show that compared with centralized learning method, the model performance is slightly lower and the training speed is accelerated to a large extent [30]; moreover, Wang Sheng Sheng et al. (2021) proposed a kind of machine learning method based on federated learning and blockchain for the data protection requirements in the model training for the segmentation of new coronary pneumonia chest CT image [31]. In multimedia field, Zhao Yu et al. (2020) proposed a kind of federated learning method based on lightweight neural network and sub-scenario model training for the problem of high latency in video surveillance, and research result show that the method could improve the accuracy and training speed compared with the benchmark method [32].

III. Application of Fusion Machine Learning

As an application of fusion machine learning in natural language text classification, this section studies the modeling of car service complaint text classification based on ensemble learning.

The research data of this section is from the Beijing Car Quality Net Information Technology Limited Company, and the dataset used includes 7 classes of car service complaint text data，the total data amount is 2100，and the data amount of every class is 300.

The technology route is shown in Fig. 2

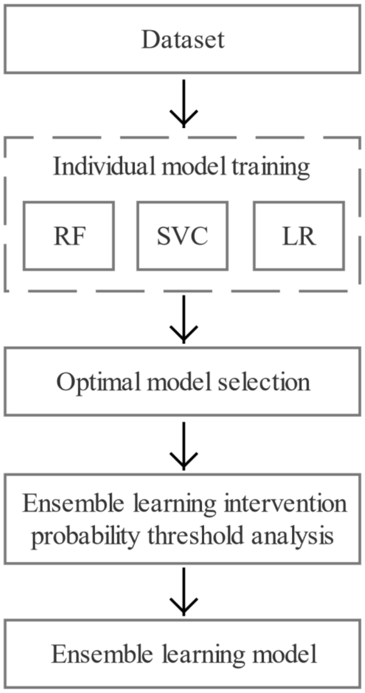

Dataset
Individual model training
RF
SVC
LR
Optimal model selection
Ensemble learning intervention probability threshold analysis
Ensemble learning model

Fig. 2. The Technical Route of Ensemble Learning.

In previous research foundation, we have conducted research on the classification modeling of car service complaint texts based on RF method and formed corresponding academic paper; this section mainly focuses on the modeling optimization based on ensemble learning. We take the RF model as a basic, and when the highest predicted probability of the RF model for different classes is less than the threshold, select the class corresponding to the highest predicted probability of the RF model, SVC model, and LR model as the output of the integrated learning model; the comparison of the model prediction effects under different threshold conditions is shown in Table III.

TABLE III. Prediction Comparison under Different Intervention Threshold Conditions

No.	Threshold	Accuracy	Recall	F1-score
1	0.1	0.8476	0.8476	0.8495
2	0.2	0.8476	0.8476	0.8495
3	0.3	0.8524	0.8524	0.8539
4	0.4	0.8571	0.8571	0.8597
5	0.5	0.8476	0.8476	0.8498
6	0.6	0.8429	0.8429	0.8447
7	0.7	0.8429	0.8429	0.8447
8	0.8	0.8381	0.8381	0.8397
9	0.9	0.8381	0.8381	0.8397
10	1.0	0.8381	0.8381	0.8397

Comparison of the prediction effects from different models is shown in Table IV. The research results show that the ensemble learning model has the optimal effect, which could effectively classify the car service complaint texts.

TABLE IV. Comparison of the Prediction Effects from different Models

No.	Method	Accuracy	Recall	F1-score
1	RF	0.8476	0.8476	0.8495
2	SVC	0.8000	0.8000	0.8022
3	LR	0.8381	0.8381	0.8408
4	Ensemble learning (Threshold: 0.4)	0.8571	0.8571	0.8597

In general, the threshold-based ensemble learning is an effective approach to improve the modeling effect of natural language text classification, and the methods based on fusion enrich the machine learning to a great extent.

IV. Conclusion and Outlook

Generally, different kinds of fusion methods have corresponding advantages, applicability and limitation, the rapid development of fusion machine learning has built a better foundation for further data science research. For the technological development trend, in terms of basic research, the innovation in the model integration strategies for ensemble learning and the innovation in the parameter fusion strategies for federated learning would be of great significance; in terms of application, the development of fusion machine learning has enriched the data modeling methods greatly, and follow-up related research could take the characteristics of specific scenario as basis to comprehensively consider the applicability of traditional machine learning methods and different types of fusion machine learning methods to achieve high-quality data modeling. In particular, the effective combination of federated learning and edge computing would make important social value in the fields of big data processing and privacy protection.

In order to provide convenient reference for relevant researchers, we have sorted up the research reports cited in this paper according to the research field, research object and fusion strategy, as shown in Table V, it is expected that subsequent researchers could make quick reference based on the reference table or take it as the basis for further improvement.

TABLE V. APPLICATION FIELDS OF THE MACHINE LEARNING BASED ON FUSION AND THE FUSION METHODS

Field	**Research object**	**Fusion**	**Reference**
Agriculture	Rice seed vigor detection	Transfer learning	[3]
	Strains classification of anoectochilus roxburghii	Ensemble learning	[12]
	Estimation of summer corn fractional vegetation coverage	Ensemble learning	[14]
	Classification of rice phenomics entitie	Ensemble learning	[15]
	Estimation of nitrogen contents in citrus leaves	Ensemble learning	[16]
	Classification of black Goji berry	Ensemble learning	[17]
	Image recognition of cotton leaf diseases and pests	Transfer learning	[18]
	Crop disease classification	Transfer learning	[21]
	Crop diseases detection	Transfer learning	[22]
	Grape leaf disease detection	Transfer learning	[24]
	Grape leaf disease recognition	Transfer learning	[25]
Energy	Power tower detection in remote sensing imagery	Transfer learning	[5]
	Electricity theft detection	Ensemble learning	[7]
Environment	Water bloom prediction	Ensemble learning	[10]
	Gaseous nitrous prediction	Ensemble learning	[13]
Medicine	Medical image object detection	Federated learning	[30]
	Chest CT image segmentation	Federated learning	[31]
Transportation	Federated learning in high-speed internet of vehicles	Federated learning	[28]
	Channel estimation	Federated learning	[29]
	Vehicular abnormal behaviors detection	Ensemble learning	[1]
Mechanical	Bearing remaining useful life prediction	Transfer learning	[4]
	Abnormal condition identification for the electro-fused magnesia	Transfer learning	[6]
	Mechanical failure warning	Ensemble learning	[2]
	Mechanical fault diagnosis	Transfer learning	[23]
Multimedia	Speech recognition	Transfer learning	[26]
	Video surveillance	Federated learning	[32]
Finance	Personal credit evaluation	Ensemble learning	[8]
Others	Character recognition	Ensemble learning	[9]
	Infrared spectroscopy data analysis	Ensemble learning	[11]
	Flotation performance recognition	Transfer learning	[19]
	Intelligent lithology identification	Transfer learning	[20]
	Wafer map defect recognition	Transfer learning	[27]

REFERENCES

[1] Xue Hong Wei, Liu Ying, Zhuang Wei Chao, Yin Guo Dong, A detection method of vehicular abnormal behaviors in V2X environment based on stacking ensemble learning[J], Automotive Engineering, 2021, 43(4): 501-508, 536.

[2] Liu Chang Liang, Wang Zi Qi, Fault early warning of wind turbine gearbox based on MSET and ensemble learning[J], Acta Energiae Solaris Sinica, 2020, 41(11): 228-233.

[3] Lu Wei, Zhang Zi Xu, Cai Miao Miao, Zhang Yi Feng, Detection of rice seeds vigor based on photoacoustic spectrum combined with TCA transfer learning[J], Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(22): 341-348.

[4] Wang Xin Gang, Han Kai Zhong, Wang Chao, Li Lin, Bearing remaining useful life prediction method based on transfer learning[J], Journal of Northeastern University(Natural Science), 2021, 42(5): 665-672.

[5] Zheng Xin, Pan Bin, Zhang Jian, Power tower detection in remote sensing imagery based on deformable network and transfer learning[J], Acta Geodaetica et Cartographica Sinica, 2020, 49(8): 1042-1050.

[6] Yan Hao, Wang Fu Li, Sun Yu Feng, He Da Kuo, Abnormal condition identification based on Bayesian network parameter transfer learning for the electro-fused magnesia[J], Acta Automatica Sinica, 2021, 47(1): 197-208.

[7] You Wen Xia, Shen Kun, Yang Nan, Li Qing Qing, Wu Yong Hua, et al., Electricity theft detection based on bagging heterogeneous ensemble learning[J], Automation of Electric Power Systems, 2021, 45(2): 105-113.

[8] Cao Jie, Shao Xiao Xiao, Analysis of personal credit evaluation method based on information gain and bagging integration learning algorithm[J], Mathematics in Practice and Theory, 2016, 46(8): 90-98.

[9] Liu Yu Xia, Lv Hong, Hu Tao, Sun Xiao Hu, Research on character recognition based on bagging ensemble learning[J], Computer Engineering and Applications, 2012, 48(33): 194-196, 211.

[10] Ma Xin Yu, Shi Yan, Wang Xiao Yi, Xu Ji Ping, Wang Li, et al. Research of water bloom prediction based on bagging ensemble learning[J], Computers and Applied Chemistry, 2014, 31(2): 140-144.

[11] Jiang Wei Wei, Lu Chang Hua, Zhang Yu Jun, Ju Wei, Wang Ji Zhou, et al., Research on a quantitative regression model of the infrared spectrum based on the integrated learning algorithm[J], Spectroscopy and Spectral Analysis, 2021, 41(4): 1119-1124.

[12] Xie Wen Yong, Chai Qin Qin, Gan Yong Hui, Chen Shu Di, Zhang Xun, et al., Strains classification of Anoectochilus roxburghii using multi-feature extraction and stacking ensemble learning[J], Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(14): 203-210.

[13] Tang Ke, Qin Min, Zhao Xing, Duan Jun, Fang Wu, et al., Prediction of gaseous nitrous acid based on stacking ensemble learning model[J], China Environmental Science, 2020, 40(2): 582-590.

[14] Zhang Hong Ming, Chen Li Jun, Liu Wen, Han Wen Ting, Zhang Shu Yin, et al., Estimation of summer corn fractional vegetation coverage based on stacking ensemble learning[J], Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(7): 195-202.

[15] Yuan Pei Sen, Yang Cheng lin, Song Yu Hong, Zhai Zhao Yu, Xu Huan Liang, Classification of rice phenomics entities based on stacking ensemble learning[J], Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(11): 144-152.

[16] Wu Tong, Li Yong, Ge Ying, Liu Ling Jie, Xi Shun Zhong, et al., Estimation of nitrogen contents in citrus leaves using stacking ensemble learning[J], Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(13): 163-171.

[17] Lu Wei, Cai Miao Miao, Zhang Qiang, Li Shan, Fast classification method of black Goji berry (Lycium Ruthenicum Murr.) based on hyperspectral and ensemble learning[J], Spectroscopy and Spectral Analysis, 2021, 41(7) : 2196-2204.

[18] Zhao Li Xin, Hou Fa Dong, Lyu Zheng Chao, Zhu Hui Chao, Ding Xiao Ling, Image recognition of cotton leaf diseases and pests based on transfer learning[J], Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(7): 184-191.

[19] Liao Yi Peng, Yang Jie Jie, Wang Zhi Gang, Wang Wei Xing, Flotation performance tecognition based on dual-modality convolutional neural network adaptive transfer learning[J], Acta Photonica Sinica, 2020, 49(10): 167-178."

[20] Xu Zhen Hao, Ma Wen, Lin Peng, Shi Heng, Liu Tong Hui, et al., Intelligent lithology identification based on transfer learning of rock images[J], Journal of Basic Science and Engineering, 2021, 29(5): 1075-1092.

[21] Wang Dong Fang, Wang Jun, Crop disease classification with transfer learning and residual networks[J], Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(4): 199-207.

[22] Yu Xiao Dong, Yang Meng Ji, Zhang Hai Qing, Li Dan, Tang Yi Qian, et al., Research and application of crop diseases detection method based on transfer learning[J], Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(10): 252-258.

[23] Shi Jie, Wu Xing, Liu Xiao Qin, Liu Tao, Mechanical fault diagnosis based on variational mode decomposition combined with deep transfer learning[J], Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(14): 129-137.

[24] Fan Xiang Peng, Xu Yan, Zhou Jian Ping, Li Zhi Lei, Peng Xuan, et al., Detection system for grape leaf diseases based on transfer learning and updated CNN[J], Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(6): 151-159.

[25] Su Shi Fang, Qiao Yan, Rao Yuan, Recognition of grape leaf diseases and mobile application based on transfer learning[J], Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(6): 151-159.

[26] Zhang Wei, Liu Chen, Fwei Hong Bo, Li Wei, Yu Jing Hu, et al., Research on automatic speech recognition based on a DL-T and transfer learning[J], Chinese Journal of Engineering, 2021, 43(3): 433-441.

[27] Shen Zong Li, Yu Jian Bo, Wafer map defect recognition based on transfer learning and deep forest[J], Journal of Zhejiang University(Engineering Science), 2020, 54(6): 1228-1239.

[28] Wang Jia Tui, Tan Guo Ping, Zhou Si Yuan, Clustered wireless federated learning algorithm in high-speed internet of vehicles scenes[J], Journal of Computer Applications, 2021, 41(6): 1546-1550

[29] Jing Xing Hong, Yin Zi Song, Cai Zhi Rong, He Shi Biao, Liao Yong, A Federated Learning Channel Estimation Algorithm for LTE-V2X[J], Telecommunication Engineering, 2021, 61(6): 681-688.

[30] Wang Sheng Sheng, Lu Shu Zhen, Cao Bin, Medical image object detection algorithm for privacy-preserving federated learning[J], Journal of Computer-Aided Design & Computer Graphics, 2021, 33(10): 1553-1562.

[31] Wang Sheng Sheng, Chen Jing Yu, Lu Yi Nan, COVID-19 chest CT image segmentation based on federated learning and blockchain[J], Journal of Jilin University(Engineering and Technology Edition), 2021, 51(6): 2164-2173.

[32] Zhao Yu, Yang Jie, Liu Miao, Sun Jin Long, Gui Guan, Federated learning based intelligent edge computing technique for video surveillance[J], Journal on Communications, 2020, 41(10): 109-115.

Detection of Premature Ventricular Contractions using 12-lead Dynamic ECG based on Squeeze-Excitation Residual Network

Duan Li[1], Tingting Sun[2], Yibai Xue[3], Yilin Xie[4], Xiaolei Chen[5], Jiaofen Nan[6]
School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan, China[1,2,4,5,6]
Department of Cardiothoracic Surgery, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China[3]

Abstract—**Premature ventricular contraction (PVC) is a very common arrhythmia that can originate in any part of the ventricle and is one of the important causes of sudden cardiac death. Timely and rapid detection of PVC on dynamic electrocardiogram (ECG) recording for patients with cardiovascular diseases is of great significance for clinical diagnosis. Furthermore, it can facilitate the planning and execution of radiofrequency ablation. But the dynamic ECGs can be easily contaminated by various noises and its morphological characteristics show significant variations for different patients. Though the deep learning methods achieved outstanding performance in ECG automatic recognition, there are still some limitations, such as overfitting, gradient disappearance or gradient explosion in deep networks. Therefore, a residual module is constructed using the squeeze-excitation method to alleviate the problems. A 20-layer squeeze-extraction residual network (SE-ResNet) containing multiple squeeze-extraction modules was designed for real-time PVC detection on 12-lead dynamic ECG. The algorithm was evaluated using the dynamic 12-lead ECGs in INCART database (168,379 heartbeats in total). The experimental results show that the test accuracy of the method proposed in this paper is 98.71%, and the specificity and sensitivity of PVC are 99.12% and 99.59%, respectively. Under the same dataset and experimental platform, the average recognition accuracy of our proposed method is increased by 0.73%, 1.55%, 2.9% and 1.65% compared with the results obtained by CNN, Inception, AlexNet and deep multilayer perceptron, respectively. The proposed scheme provides a new method for real-time detection of PVC on dynamic 12-lead ECGs. The experiment results show that the proposed method outperforms state-of-the-art methods, and has good potential for clinical applications.**

Keywords—Dynamic ECG; squeeze-excitation; residual network; premature ventricular contraction

I. Introduction

Premature ventricular contractions (PVCs) are the most common type of arrhythmia and, under certain conditions, can lead to life-threatening heart disease. Electrocardiogram (ECG) is not only a noninvasive and economical tool for routine cardiovascular examination, but also an essential monitoring device in surgical procedures and intensive care units. It is more clinically significant for the diagnosis of PVC. However, it is time-consuming and arduous for cardiologists to analyze many long-term dynamic ECG. Therefore, the automatic detection of PVC on body surface dynamic ECGs can not only improve cardiology workflow efficiency and timely prevent cardiac diseases such as arrhythmia, but also accurately locate the occurrence time and source localization of ventricular premature beats, and then guiding the surgical process such as radiofrequency ablation, etc.

Dynamic ECG is susceptible to various background noises, such as power-line interference, inotropic noise, baseline drift and motion artifact, and its morphological characteristics show significant variations for different patients and under different temporal and physical conditions. Even experienced specialists cannot accurately determine the type of arrhythmias. Machine learning methods such as deep neural networks can more accurately detect arrhythmias such as PVC, and have shown good clinical applications [1].

Deep learning has experienced great breakthroughs in the past decade in many fields, such as image recognition and natural language processing. With the popularity of deep learning and outstanding performance in other fields, researches use deep learning to monitor arrhythmias such as PVCs [2-5]. They transform ECG as one-dimensional time-series signals or two-dimensional signals such as multi-lead ECG beats or time-frequency images as the input of the convolution neural network (CNN). Then conduct layer-by-layer feature extraction and classification. In 2017, Acharya et al.[6] proposed a 9-layer deep CNN to discriminate 5 types of ECG heartbeats and achieved 94.03% accuracy using MIT-BIH arrhythmia database. In 2018, Yildirim et al. [7] designed a new 1D convolutional neural network model (1D-CNN) to recognize 17 different types of long-time dynamic ECG signals. Using the MIT-BIH database, they achieved an overall accuracy of 91.33% for the 17 type arrhythmias. In 2019, Andersen et al. [8] proposed a convolution combined cyclic convolution model, which could search for atrial fibrillation heartbeats from 24-hour dynamic ECG signals, and achieved good results. In 2020, Ullah et al. [9] proposed a two-dimensional (2-D) CNN model to recognize eight types of ECG signals. The model was evaluated on the MIT-BIH dataset and the classification accuracy reached 99.11%. With the deepening of deep network structure, the accuracy of the neural network model will decrease. That is the degradation of neural network. To overcome the problem, deep CNN model with residual structure is developed for ECG arrhythmias detection, which improves classification accuracy [10]. In 2019, Brito et al. [11] proposed a deep learning model based on ResNet architecture. They conducted experiments using MIT-

BIH arrhythmia database and achieved an accuracy of more than 90%. In 2020, Li et al. [12] classified arrhythmias based on deep residual network. The experiments applied to the MIT-BIH arrhythmia database and showed high classification performance with an accuracy of 99.38%. Deep learning has made some achievements in the classification and recognition of PVCs [13, 14]. It has good performance on small evaluation samples and static ECGs, but the accuracy decreases for clinical dynamic 12-lead large data especially for non-equilibrium dataset. Therefore, the evolutionary model and method is crucial to improve the efficiency and effects of PVC detection which will promote the further clinical applications.

Although there are many studies on arrhythmia detection in the literature, there are still various problems such as difficult convergence of deep networks, training cost, and computational complexity. Furthermore, in the literature, most models are trained on relatively clean open-source ECG datasets such as the MIT-BIH database. In this study, considering the advantages and disadvantages of existing technologies, a squeeze-excitation module is constructed which embedded in a residual structure to improve the convergence of the deep network. It aims to improve the non-linear fitting ability of the deep network by reconstructing the hyperplane parameters through the squeeze-excitation operation. The network model uses a feature rescaling strategy, where the importance of each feature channel is automatically obtained by learning, and then the useful features are promoted and the less useful features are suppressed according to its importance. The network model can fully consider the weight of each lead and main wave of ECG signals, and provide a new idea for deep feature extraction of arrhythmia heartbeats. Based on the SE-ResNet model, the performance of the model was evaluated by 168379 12-lead heartbeats from the St Petersburg INCART 12-Lead (INCART) arrhythmia dynamic ECG database. The effectiveness of this method is evaluated by experiments.

The remainder of this paper is organized as follows. Section II described the related work and methods for this study. In Section III, a novel SE-ResNet for detection of premature ventricular beats was implementation. Experimental results are also described. Section IV is discussion and Section V concludes the paper.

II. Method

A. Convolutional Neural Network

CNN is a deep learning algorithm based on artificial neural network structure, trained by a gradient-based optimization algorithm. In contrast to traditional machine learning algorithms, CNN architecture does not need to manually extract features from raw data. Feature extraction and classification are embedded in the architecture, so robust features can be automatically identified from the input data [15]. In general, a CNN consists of multiple back-to-back layers connected in a feedforward manner. As shown in Fig. 1, in the CNN architecture, the main layers are including convolutional layers, pooling layers and a fully-connected layer. Convolutional and pooling layers are responsible for feature extraction, while fully-connected layer is responsible for classification.

B. Squeeze-excitation Residual Network

Bioelectric signals are characterized by individual variability, strong interference, and multi-lead characteristics. Individual variability is reflected in the ECG morphology of different patients with the same disease, and even the difference and translation of characteristic wave directions. In addition, the same patient will also have certain differences in different times and environments. Different leads of the ECG signal reflect the potential transformation of cardiac activity in different parts of the body. The waveforms corresponding to each lead has great difference, and each lead is relatively independent. CNN improves performance by deepening the network structure as much as possible. However, with the increase of CNN depth, namely, the number of network layer increases, the performance of the model tends to saturate and even decline rapidly, which makes the training of deep networks more difficult.

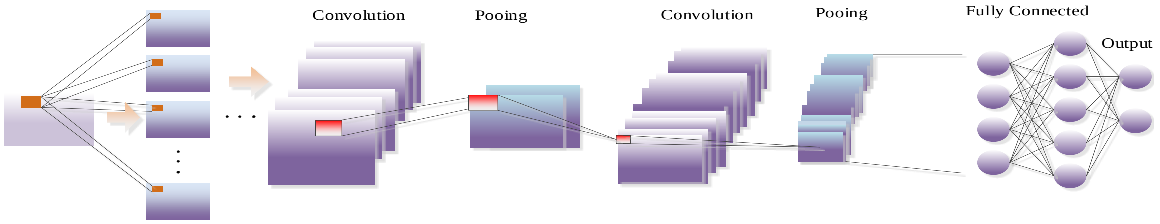


Fig. 1. CNN Architecture.

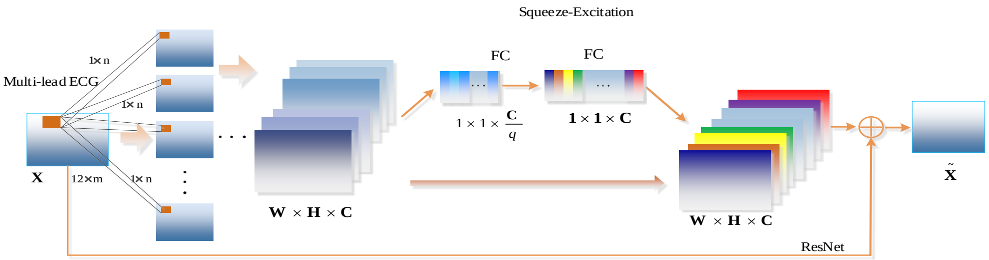


Fig. 2. SE-ResNet Architecture.

Aiming at the above problems, this paper adopts the SE-ResNet model, a deep network architecture with stronger nonlinear fitting ability. The model constructs a squeeze-excitation module which embedded in a residual structure. The network model adopts a feature recalculation strategy, namely, the importance of each feature channel is automatically obtained by learning, and then the useful features are improved and the features that are not useful for the current task are suppressed according to the importance. As shown in Fig. 2, in the squeeze layer, for inputs X = [x_1,x_2, ⋯ ,x_c], where $\mathrm{x_c} \in R^{H\times W}$, the simplest aggregation technique, global averaging, is used to generate channel statistics. Formally, the statistic $z \in R^{c}$ is generated by reducing X by reducing its spatial dimension H × W, and the c-th element of z is calculated by the following equation (1):

$$\mathrm{z_c} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{i=1}^{W}\mathrm{x}_c(i,j) \tag{1}$$

Among them, the H and W represent the height and width of the feature map, respectively, and C represents the number of feature map channels.

In the excitation layer, two fully-connected layers (FC) are used to achieve channel scaling with a reduction rate of q. The dimension of feature data changes from 1 × C to 1 × C / q, and then playback to 1 × C. Finally, the sigmoid activation function is used to scale the data back to the previous data dimension. Since we want to ensure that multiple channels are allowed to be emphasized, a simple gating mechanism with sigmoid activation in equation (2) is employed:

$$s = \sigma(g(z,W)) = \sigma(W_2\sigma(W_1 z)) \tag{2}$$

where σ is the sigmoid function, $W_1 \in R^{\frac{c}{q}\times C}$ and $W_2 \in R^{C\times\frac{\mathrm{c}}{q}}$. Here q is a scaling parameter. The final output of the block is obtained by rescaling X with the activations s, as shown in Equation (3) below:

$$\tilde{\mathrm{x}}_c = s_c \times \mathrm{x}_c \tag{3}$$

Where $\tilde{X}$ = [$\tilde{\mathrm{x}}_1$,$\tilde{}$ $\tilde{\mathrm{x}}_2$,…,$\tilde{}$ $\tilde{\mathrm{x}}_c$] and the channel multiplication between the index quantity sc and the feature map $\mathrm{x_c} \in R^{H\times W}$.

Reconstructing hyperplane parameters by squeeze-excitation operation can alleviate the problems of the difficulty tuning of deep network and the nonlinear fitting ability of deep network. The architecture can effectively avoid the effect of traditional ECG signal feature extraction on the subsequent classification accuracy.

III. Detection of Premature Ventricular Contractions based on SE-ResNet

A. Data Sources and Evaluation Metrics

In this experiment, the 12-lead dynamic ECGs from the open INCART Arrhythmia database were used for evaluating the algorithm. The INCART Dynamic Arrhythmia ECG database contains 75 records, sampled at 275 Hz. Each record is about half an hour long and has 12 leads. The original ECG data were collected from patients who were examined for coronary artery disease, and most of them had premature ventricular contractions [16].

Since V6 lead is missing in 102 ECG record, V3 lead is missing in 103 record, and V4 lead is missing in 158 record, the above three ECG records were deleted in consideration of the lead consistency. All heartbeats from the remaining 72 records were used in this experiment. According to our statistics, the data in INCART database included 168379 heartbeats. In order to test the recognition effect of premature ventricular contractions, all cardiac heartbeats were divided into three types: normal heartbeats (N), premature ventricular contractions (V) and other heartbeats (O). The number of normal heartbeats was 143,260, the number of premature ventricular contractions was 19,640 and the number of other heartbeats (premature atrial contractions, supraventricular premature heartbeats and right bundle branch block, etc.) was 5479.

In this study, three metrics were used to evaluate the performance of the proposed classification method: accuracy ($A\mathrm{cc}$), sensitivity ($S\mathrm{e}$) and specificity (Sp), which were defined in formulas (4), (5) and (6) respectively. The calculations are made based on the statistical results of multiple experiments.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$S\mathrm{e} = \frac{TP}{TP+FN} \tag{5}$$

$$S\mathrm{p} = \frac{TN}{TN+FP} \tag{6}$$

Among them, TP, FP, TN and FN are true positive, false positive, true negative and false negative, respectively [17, 18].

B. Pre-processing

In contrast, dynamic ECG signals have stronger interference from noise such as motion artefacts, due to factors such as poor skin-to-electrode contact, the effects of breathing and poor contact with the power lines of electrical equipment, et al. Therefore the signal characteristics are chaotic, non-linear and multi-channel. Noises in the ECG signal distort some of its morphological characteristics, which make diagnosis more difficult. So reasonable filtering is very important for subsequent recognition. The useful part of the frequency in the ECG signal mainly ranges from 1 to 40 Hz, and the interest signal is easily submerged in the background noises. The main noise sources include baseline drift, power frequency interference, motion artifact and myoelectric interference [19]. In addition, the low-frequency part of ECG signal contains indicators of malignant arrhythmias such as S-T segment abnormalities, and the high frequency part reflects the amplitude information of main complex wave. In order to keep the morphological characteristics of ECG signal as completely as possible, a wavelet adaptive threshold filtering method is proposed. The algorithm includes three steps: wavelet decomposition, adaptive threshold de-noising and reconstruction. The selection of threshold is adaptive to the signal; therefore, the inherent morphological characteristics of the ECG signal are preserved as much as possible.

The INCART dynamic ECG database contains long-term ECG records with complete annotation of heartbeats. Before applying the model, segmentation of the heartbeat was performed, dividing the long ECG record into heartbeat segments that represent different types, namely, N, V and O. Each heartbeat segment was extracted by selecting a window of 300 samples around the R-peak, which formed by taking 92 points in front of the R-peak and 137 points behind the R-peak, respectively. As each heartbeat segment consists of 230 samples, and the ECG signals are 12 leads, so the size of each beat sample is 12 * 230.

C. SE-ResNet Network Modeling

The architecture of the SE-ResNet network pay more attention to the weights of each lead, and can fully extract the morphological features of the multi-lead ECG waveform. Since the number and complexity of network layers will have a great impact on the training results, we designed SE-ResNet models with different layers and structures, and performed several cross-validation experiments and comparisons. Fig. 3 is the experimental results of the accuracy for each epoch training using the SE-ResNet networks with 8, 12, 16, 20, and 24 layers. As shown in Fig. 3, the 20-layer SE-ResNet achieved better training and testing results, and the 24-layer network performance was comparable to the 20-layer network. Considering computation efficiency and the real-time implementation, especially performed on the embedded processor, we select a 20-layer network for PVC recognition in the following experiments.

In order to optimize the model, we selected the Sgd optimizer, with an initial learning rate of 0.05. To improve the performance of the neural networks, the negative log-likelihood loss (NLLLoss) function and cross-entropy loss (CrossEntropyLoss) function were compared, and the results were shown in Fig. 4. The figure represents the convergence performance for 30 epoch training iterations using different loss functions. As is shown in Fig. 4, using the CrossEntropy loss function, the neural network has more robust stability and better convergence of the training process. Therefore, the CrossEntropyLoss function was selected as the loss function in the following experiments.

In this work, the model architecture of the 20-layer SE-ResNet was designed, as is shown in Fig. 5. The main difference between the proposed network and the original SE-ResNet on ImageNet is that the proposed network uses 1D convolutions instead of 2D convolutions. The input ECG heartbeat is a sample of 12*230 size, and after a convolutional layer with a kernel size of 1*3, followed 9 residual blocks. The residual blocks of the network have 16, 32, and 64 channels respectively. Among them, the first three residual blocks have 16 channels, the second three residual blocks have 32 channels and the remainder three residual blocks has 64 channels.

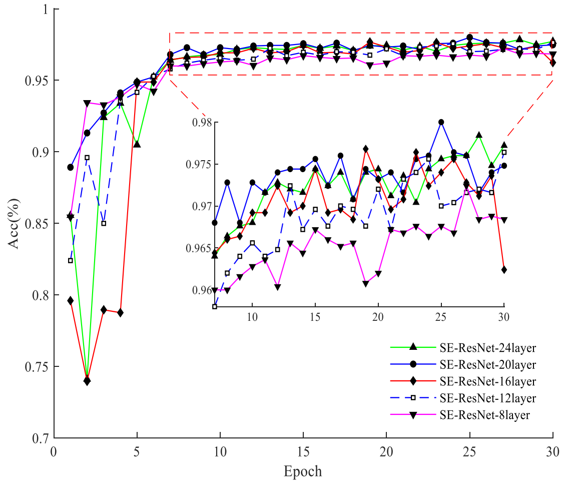


Fig. 3. ECG Recognition Results of different Network Architecture.

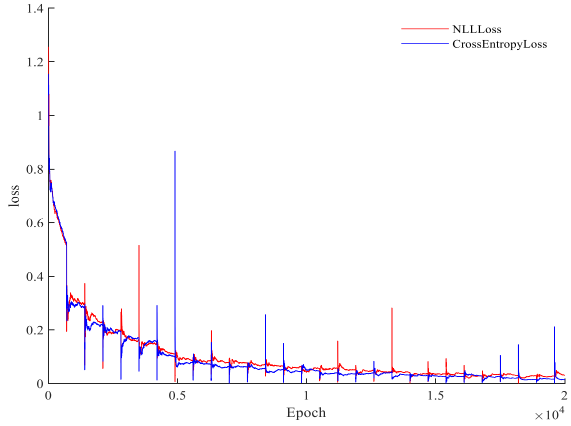


Fig. 4. Comparison of Convergence in Training Process with different Loss Functions.

Each residual block is embedded with squeeze-excitation modules, so a total of 3+ 3+3=9 residual blocks are designed in Fig. 5. In order to accurately extract information from each channel of the ECG heartbeat signal, each residual (short-connect connection) is embedded with a squeeze-excitation module that automatically captures the weight information of each feature channel of the ECG heartbeats in a learning manner, effectively enhancing the features of the useful channels and suppressing information that is not sufficiently useful for the current classification recognition task. Take the first three residual blocks (short-connect) of 16 channels as an example, as shown in Fig. 5, in order to make the best use of the contextual ECG feature information of each channel, channel-level statistics are generated by global averaging pooling. The excitation layer is implemented using two fully-connected layers (FC) for channel scaling, the reduction rate is taken as 4, the dimension of the feature data is changed from 1* 16 to 1* 4 and then replayed as 1* 16. Finally, use the sigmoid activation function to rescale the data back to the dimensions before squeezing. It is equivalent to map the data associated with the input to a set of channel weights, so that the channel features are not limited to the local perceptual field of the convolutional network. Therefore, the context information can be easily understood, and different weights are assigned to the channels. Residual architecture can improve the parameter adjustment ability of the network, namely, the optimization ability. As shown in Fig. 5, there are three 16-channel SE-ResNet modules in total, and each module has two layers of convolution. There are three 32-channel SE-ResNet modules in total, and each module has two layers of convolution. Similarly, there are three 64-channel SE-ResNet modules in total, each with two layers of convolution. Through 19 layers of network transmission, deep feature extraction is completed. Finally, the feature is fed into a fully-connected layer, so the network model has a total of 20 parameter layers.

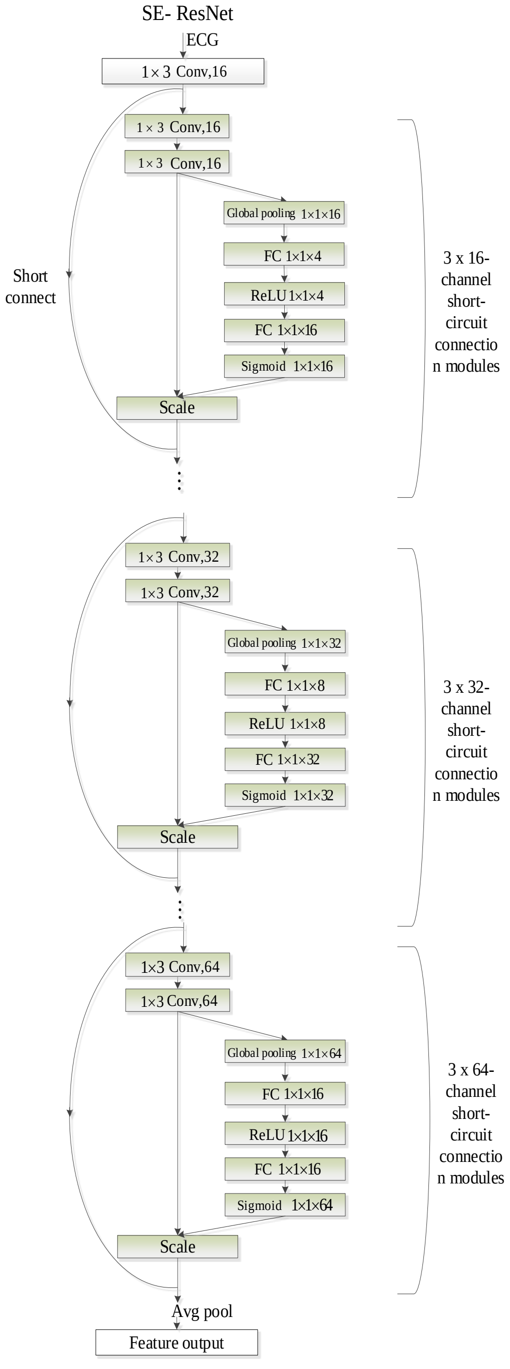


Fig. 5. SE-ResNet Network Model.

D. Analysis of Experimental Results

In order to evaluate the recognition effect of the proposed SE-ResNet model for PVCs recognition based on the INCART dynamic ECG database, this experiment employed a wavelet self-adapting filter to preprocess all 72 lead-consistent ECG records, and segmented 168379 heartbeats according to the R-peak position. We randomly selected 3000 normal heartbeats, 3000 PVCs and 1000 other type heartbeats, respectively from the total 168,379 heartbeats which include 143260 normal heartbeats, 19640 premature ventricular contraction and 5479 other heartbeats. So the total training samples is 7000. Similarly, 1000, 1000 and 500 heartbeats were randomly selected from the remainder data, which were used for cross verification. The remainder 158,879 heartbeats were used for the final evaluation. The classification results were expected to be affected by the heartbeat sample size, and the ratio of training to testing sample size. In general, the lower the ratio of training samples to total samples, the greater the generalization ability of the classifiers. In order to evaluate the network's performance and avoid the occasionality of random sampling, we carried out three random sampling experiments, and the confusion matrix was calculated from the average value of the three experiment results. Then, CNN, Inception, multi-layer perceptron (MLP) and Alexnet of the same complexity were designed. In these experiments, the parameters of the networks were regulated to be best fitted in the classification task. The five networks were trained 30 epochs using the same training samples, and the test results were also compared. The confusion matrix and evaluation metrics are shown in Table I.

Since the experiments were performed three times, three trained models were obtained with exactly the same training dataset. When testing the 12-lead ECG signal, the predicted probabilities of the three models were averaged and finally got the final classification results. The confusion matrix and three different statistical indices, such as Sensitivity (Se), Specificity (Sp), and Accuracy (Acc) were summarized in Table I. As is shown in Table I, it can be seen that the algorithm designed in this paper achieved an accuracy of 98.71% for the detection of the arrhythmias, and the sensitivity (Se) and specificity (Sp) for PVC recognition are 99.12% and 99.59%, respectively. The experimental results showed that the accuracy of the designed SE-ResNet algorithm was improved by 0.73%, 1.55%, 1.65% and 2.9% over CNN, Inception, MLP and AlexNet networks, respectively. The PVC sensitivity (Se) value of the SE-ResNet test was 98.71%, and the sensitivities (Se) of the CNN, Inception, MLP, and Alexnet tests were 92.17%, 97.74%, 94.62, and 95.97%, respectively. It can be seen that the sensitivity of PVC has been significantly improved.

E. Compare Results with other Methods

The PVCs detection results of our proposed method were compared with the recent published research results. These published results on the same dataset are shown in Table II. The detection results were expected to be affected by the heartbeat sample size, the number of classes for classification, and the ratio of training to testing sample size. In general, the lower the ratio of training samples to total samples, the greater the generalization ability of the classifiers, and the fewer the types for classification, the higher the recognition accuracy. In one study, Al Rahhal et al. [20] proposed an electrocardiogram

(ECG) technique based on multi-lead signals and a deep learning architecture. Automatic identification of ECG signals was performed using INCART Arrhythmia Database, which automatically recognized three types: normal heartbeats, PVC and other heartbeats. The overall classification accuracy reached 98.6%, but the sensitivity (Se) of PVC was 91.4%. In a study, Allami [21] used three morphological features and seven statistical features, and also employed the artificial neural network (ANN) classifier for PVC and non-PVC ECG heartbeats recognition. Using 75 ECG records, the classification accuracy and the sensitivity (Se) of PVC they achieved was 95.8% and 93.9%, respectively. In another, Malek et al. [22] developed an improved template matching technique for PVCs and normal heartbeats detecting, by analyzing the maximum value and the correlation coefficients of the maximum and minimum value. The classification accuracy rate was 97.91%, and the sensitivity (Se) and specificity (Sp) of PVC detection were 91.14% and 98.82%, respectively. In general, the proposed SE-ResNet residual network has achieved superior performance on the PVCs recognition experiments using the 12-lead dynamic ECGs. It demonstrates great clinical application prospects.

F. Discussion

Dynamic 12-lead ECG is the gold standard in the detection of arrhythmias. Multi-lead dynamic ECG has strong background noise; there are correlations of ECG leads. Different arrhythmias have corresponding lead characteristics, and the main wave morphological characteristics of some leads are more distinguishable for the specific arrhythmia. Squeeze-excitation network introduces attention mechanism that the importance of each feature channel is automatically obtained by learning. That the useful features are promoted and the less useful features are suppressed according to its importance. The network model can fully consider the weight of each lead and main wave of ECG signals. Therefore, the model can fully extract the morphological features of multi-lead and its main waves, and improve the robustness and generalization ability of the network.

TABLE I. COMPARISON OF PVC RECOGNITION RESULTS ON INCART DATABASE

Method	Confusion Matrix				Evaluation		
		N	***V***	***T***	***Se***	***Sp***	***Acc***
SE-ResNet	N	137682	496	1082	98.87%	98.22%	98.71%
	V	97	15503	40	99.12%	99.59%	
	T	252	87	3640	91.48%	99.28%	
CNN	N	138079	759	422	99.15%	91.70%	97.98%
	V	862	14416	362	92.17%	99.44%	
	T	766	44	3169	79.64%	99.49%	
Inception	N	135780	1332	2148	97.50%	96.41%	97.16%
	V	142	15287	211	97.74%	98.99%	
	T	563	120	3296	82.83%	98.48%	
MLP	N	136246	1255	1759	97.84%	95.65%	97.06%
	V	204	14798	638	94.62%	99.00%	
	T	650	171	3158	79.37%	98.45%	
Alexnet	N	134939	4202	119	96.90%	88.64%	95.81%
	V	624	15010	6	95.97%	96.99%	
	T	1605	103	2271	57.07%	99.92%	

TABLE II. COMPARISON WITH OTHER METHODS PROPOSED IN THE LITERATURES

Method	Classification	Se	Sp	Acc
Al Rahhal et al.[20]	3	91.4%	*	98.6%
Allami[21]	2	93.9%	*	95.8%
Malek et al.[22]	2	91.14%	98.82%	97.91%
Algoritm of this paper	3	99.12%	99.59%	98.71%

In this work, we analysis the characteristics of the long-time dynamic 12-lead ECGs and introduce the squeeze-excitation ResNet model to the real-time PVCs recognition on 12-lead dynamic ECGs. Which is to overcome the and fully extract the multi-lead and multi-dominant complexes. Reconstructing hyperplane parameters by squeeze-excitation operation can alleviate the problems of the difficulty tuning of deep network and the nonlinear fitting ability of deep network. In addition, we considered the influence of SE-ResNet models with different layers and structures on the training results, and through experiments, a 20-layer network was selected as the recognition model. At the same time, the results of two different loss functions under the 20-layer network model are also compared. Finally, the CrossEntropyLoss function was selected as the loss function. The test accuracy of the method proposed in this paper is 98.71%. Under the same dataset and experimental platform, the recognition accuracy of this method is improved compared with CNN, IncexNet and deep multilayer perceptron. The proposed SE-ResNet residual network has achieved superior performance on the PVCs recognition experiments using the 12-lead dynamic ECGs. It demonstrates great clinical application prospects. Then, the PVCs detection results of our proposed method were compared with the recent published research results. Experimental results show that our method has better precision and accuracy than previous studies. Which perform Demonstrate its practical application potential in the medical field.

Although the research results of this paper has achieved good performance, there are still some challenges and study values. First, Bioelectric signals are characterized by individual variability, strong interference, and multi-lead characteristics. The same patient will also have certain differences in different times and environments. Real-time identification of multiple arrhythmias using clinical big data is a challenge work and is of great value in clinical diagnosis. Therefore, we will extend the proposed method to multiple types of arrhythmias on long-term dynamic ECGs, which is of great significance for collecting more clinical ECG data from different patients under different conditions. This also puts forward higher requirements on the robustness and generalization ability of the recognition network. Second, because of the fast and slow changes in heart rhythm, it is not always desirable to use a fixed beat length. It is necessary to study adaptive beat size segmentation to meet different. Third, to achieve higher accuracy, many studies focus on the deep learning trend of making networks deeper and more complex. However, many real-world studies must be performed on computationally limited platforms. We need to consider the computational speed and the computational complexity of the model, as well as its accuracy. Finally, as catheter ablation is an effective therapy for treatment of symptomatic PVCs. And it is important to estimate the targeted anatomic ablation site that prior to the procedure. Based on this study, the localization of the site of origin of a PVCs using 12-lead ECGs is still an interesting and challenge work. Therefore, the further study of us will focus on the location of PVCs. It is important for the planning and execution of the electrophysiological procedure for the catheter ablation and has great clinical application values.

IV. Conclusion

In this study, a 12-lead dynamic ECG PVCs recognition algorithm based on squeeze-excitation residual network is proposed. The squeeze-excitation module is constructed and embedded in the residual structure to improve the performance of the deep network. The hyperplane parameters are reconstructed by squeezing-excitation operations to improve the nonlinear fitting ability of deep networks. A SE-ResNet model based on 20 layers is designed, which overcomes the degradation problem caused by the increase of the network layers when the deep neural network approximates the identity mapping, and ensures the smooth convergence of the network. Experiments of PVCs recognition was performed using 168,379 heartbeats from the INCART dynamic 12-lead ECG database. In the same experimental samples, several popular deep neural network algorithms were compared. The experimental results show that the proposed method effectively improves the overall PVC recognition accuracy on the INCART 12-lead dynamic ECGs, as well as the sensitivity and the specificity have achieved. In the future, we intend to improve the performance of this work with further advanced deep learning techniques. Additional datasets will be added to test the performance of the model to further verify the robustness of the method used. Furthermore, we will further study the localization of the site of origin of a PVC, which is important for the planning and execution of the electrophysiological procedure for the catheter ablation.

Acknowledgment

This work was financially supported by Henan Provincial Science and Technology Research Project (222102210219); Zhengzhou University of Light Industry Doctoral Fund (2018BSJJ031); Henan Provincial University Key Research Project (20A510014).

References

[1] Z. Zhao, X. Wang, Z. Cai, J. Li and C. Liu, "PVC Recognition for Wearable ECGs Using Modified Frequency Slice Wavelet Transform and Convolutional Neural Network," 2019 Computing in Cardiology (CinC), 2019, pp. 1-4, doi: 10.23919/CinC49843.2019.9005872.

[2] S. Hong , Y. Zhou, J. Shang, C. Xiao and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: a systematic review," Comput Biol Med, vol. 122, pp. 103801, 2020, doi: 10.1016/j.compbiomed.2020.103801.

[3] J. Huang, B. Chen, B. Yao and W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," in IEEE Access, vol. 7, pp. 92871-92880, 2019, doi: 10.1109/ACCESS.2019.2928017.

[4] M. Sraitih , Y. Jabrane and A. Hajjam El Hassani, "An Automated System for ECG Arrhythmia Detection Using Machine Learning Techniques," Journal of Clinical Medicine, vol. 10, no. 22, pp. 5450, 2021, doi: 10.3390/jcm10225450.

[5] S. Kiranyaz, O. Avci; O. Abdeljaber, T. Ince, M. Gabbouj and D. J. Inman, "1D convolutional neural networks and applications: A survey," Mech Syst Signal Process, vol. 151, pp. 107398, 2021, doi: 10.1016/j.ymssp.2020.107398.

[6] U. R. Acharya , S. L. Oh , Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," Computers in Biology and Medicine, vol. 89, pp. 389-396, 2017, doi: 10.1016/j.compbiomed.2017.08.022.

[7] O. Yildirim, P. Plawiak , R. S. Tan and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," Computers in Biology and Medicine, vol. 102, pp. 411-420, 2018, doi: 10.1016/j.compbiomed.2018.09.009.

[8] R. S. Andersen, A. Peimankar and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," Expert Systems With Applications, vol. 115, pp. 465-473, 2019, doi: 10.1016/j.eswa.2018.08.011.

[9] A. Ullah, S. M. Anwar and R. M. Mehmood, "Classification of Arrhythmia by Using Deep Learning with 2-D ECG Spectral Image Representation," Remote Sensing, vol. 12, no. 10 pp. 1685, 2020, doi: 10.3390/rs12101685.

[10] P. Rajpurkar , A. Y. Hannun , M. Haghpanahi, C. Bourn and A. Y. Ng, "Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks," ArXiv, 2017.

[11] C. Brito, A. Machado and A. Sousa, "Electrocardiogram beat-classification based on a ResNet network," Stud Health Technol Inform, vol. 264, pp. 55-59, 2019, doi: 10.3233/SHTI190182.

[12] Z. Li , D. S. Zhou , L. Wan and W. Mou, "Heartbeat classification using deep residual convolutional neural network from 2-lead electrocardiogram," Journal of Electrocardiology, vol. 58, pp. 105-112, 2020, doi: 10.1016/j.jelectrocard.2019.11.046.

[13] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia and A. Y. Ng , "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network". Nat Med, vol. 25, pp. 65-69, 2019, doi: 10.1038/s41591-018-0268-3.

[14] B. Hou, J. Yang, P. Wang and R. Yan, "LSTM-Based Auto-Encoder Model for ECG Arrhythmias Classification," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 4, pp. 1232-1240, April 2020, doi: 10.1109/TIM.2019.2910342.

[15] Ö. Yıldırım, P. Pławiak, R. S. Tan and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," Comput Biol Med, vol. 102, pp. 411-420, 2018, doi: 10.1016/j.compbiomed.2018.09.009.

[16] A. L. Goldberger , L. A. Amaral , L. Glass , J. M. Hausdorff, P. C. h. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng and H. E. Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," Journal of the American Heart Association, vol. 101, no. 23, pp. 215-220, 2000.

[17] D. Li, R. Shi, N. Yao, F. Zhu and K. Wang. "Real-Time Patient-Specific ECG Arrhythmia Detection by Quantum Genetic Algorithm of Least Squares Twin SVM." Journal of Beijing Institute of Technology, vol. 29, no. 1, pp. 29-37, 2020, doi: 10.15918/j.jbit1004-0579.18156.

[18] L. P. Jin, J. Dong, "Deep Learning Algorithm for Clinical ECG Analysis," Chinese Science: Information Science, vol. 45, no. 3, pp. 398-416, 2015.

[19] T. Tuncer , S. Dogan , P. Plawiak and U. Rajendra Acharya, "Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals," Knowledge-Based Systems, vol. 186, pp. 104932, 2019, doi: 10.1016/j.knosys. 2019.104923.

[20] M. M. A. Rahhal, N. A. Ajlan, Y. Bazi, H. A. Hichri and T. Rabczuk, "Automatic Premature Ventricular Contractions Detection for Multi-Lead Electrocardiogram Signal," 2018 IEEE International Conference on Electro/Information Technology (EIT), 2018, pp. 0169-0173, doi: 10.1109/EIT.2018.8500197.

[21] R. Allami, "Premature ventricular contraction analysis for real-time patient monitoring," Biomedical Signal Processing and Control, vol. 47, pp. 358-365, 2019, doi: 10.1016/j.bspc.2018.08.040.

[22] A. S. Malek, A. Elnahrawy, H. Anwar and M. Naeem, "Automated detection of premature ventricular contraction in ECG signals using enhanced template matching algorithm," Biomedical Physics & Engineering Express, vol. 6, pp. 15-24, 2020, doi: 10.1088/2057-1976/ab6995.

Personality Classification Model of Social Network Profiles based on their Activities and Contents

Mervat Ragab Bakry[1], Mona Mohamed Nasr[2], Fahad Kamal Alsheref[3]
Information Systems Department, Beni-Suef University, Egypt[1, 3]
Information Systems Department, Helwan University, Egypt[2]

Abstract*—Social networks have become an important part of everyday life, especially after the latest technologies such as smartphones, tablets, and laptops have become widespread. Individuals spend a lot of time on social media and express their feelings and opinions through statuses, comments, and updates which could be a way to understand and classify their personalities. The personalities in psychological science are divided into five classes according to the Big-five model (Openness, Extraversion, Consciousness, Agreeableness, and Neurotic). This model shows the key features with their weights for each personality. In this paper, a proposed model is developed for detecting the personality features from users' activities in social networks. In this model, machine learning techniques are used for predicting the personalities with a score for each Big-five model type and sorting them in descending order. The personality classification model will be useful in developing a better understanding of the user profile and specifically targeting users with appropriate advertising. Any social media network user's personality can be predicted by using their posts and status updates to get better accuracy. The experimental results of the model in this study provide an enhancement because it can predict the precise score of one user in each factor of the Big-five. The proposed model was tested on a dataset extracted from Facebook and manually classified by experts, and it achieved 89.37% accuracy.

***Keywords*—Psychological personality; machine learning techniques; big-five; LinearSVC**

I. Introduction

In recent years, social media has surpassed email as the most extensively utilized method for communication and engagement between people [1]. As people increasingly choose to connect informally through smartphones, face-to-face engagement is becoming less common. As a result, determining a person's personality is challenging. However, because individuals spend a lot of time on social media and express their feelings and opinions through statuses, comments, and updates, what is published on social networks can help to receive the information needed for this study.

Users on Facebook typically express their views and ideas through status updates or comments. Despite the fact that Facebook is now more extensively used to post photographs and videos, this study focuses on the linguistic element of Facebook users' status updates. There is a link between a person's personality and his or her language conduct, according to studies in the field of psychology [2] and [3]. Using a natural language processing technique, this association may be successfully evaluated and shown. As a result, the purpose of this study is to develop a system that can automatically estimate a user's personality based on their Facebook behaviors.

Big Five Personality, MBTI (Myers-Briggs Type Indicator), and DISC are some of the personality models used to predict personality (Dominance Influence Steadiness Conscientiousness). However, after some thought and a literature search, Big-five personality was chosen for this paper since it is the most common and accurate method of determining a person's personality qualities. Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism are the traits in this paradigm.

The corpus used in this study consists of a one dataset which contains 100 Facebook profiles with their 27,182 posts classified over 38 categories. The rest of this paper is organized as follows. Section II gives background information about used concepts and techniques. Section III discusses related work. Section IV describes the proposed model for personality classification model. In Section V, the case study and its findings are discussed and the conclusion in Section VI.

II. Background

Several methodologies, such as the Personality model and other classification algorithms, are employed in the proposed model, as discussed in the following sections.

A. Personality Models

Personality is a persistent tendency and attribute that governs how an individual's psychological behavior is similar and different (thoughts, emotions, behaviors). In other words, personality reflects patterns of human behavior, thought, and interpersonal communication, and has a significant impact on a variety of elements of life, including happiness, taste, and physical and mental health [3] and [4]. There are other psychological personality models, but the Big Five (shown in Fig. 1) is the most often utilized.

1) Big-five model: Many academics nowadays believe that there are five basic personality traits [5]. For several years, evidence of this notion has been growing, beginning with D. W. Fiske's (1949) research and later elaborated upon by other researchers such as Norman (1967), Smith (1967), Goldberg (1981), and McCrae & Costa (1983). (1987).

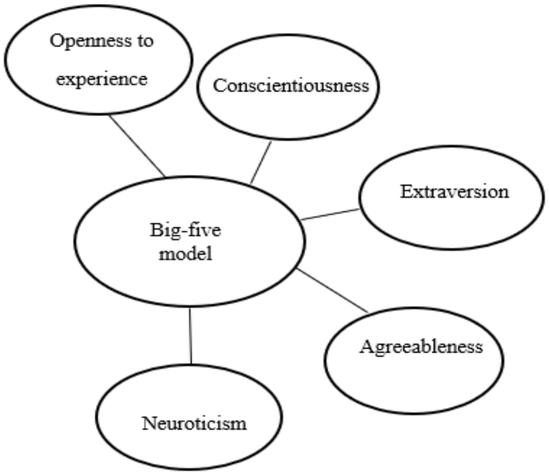


Fig. 1. Big-Five Model [5].

The big five personality traits are broad categories of personality traits are:

- Openness: Imagination and insight are two features of this attribute [5]. People with a high level of this attribute also have a diverse set of interests. They are curious in the world and other people, and they are eager to learn new skills and have a better understanding of new concerns. People with a high level of this characteristic are more likely to be daring and creative. People with low levels of this component are more traditional and may have difficulty thinking abstractly [6].
- Conscientiousness: High levels of thinking, effective impulse control, and goal-directed activities are all common aspects of this component [5]. People that are very conscientious have a propensity to be structured and detail oriented. They consider the consequences of their activities and are mindful of deadlines [6].
- Extraversion: Excitability, sociability, talkativeness, decisiveness, and extensive use of emotional emotions are all characteristics of extraversion [5]. People with a high level of extraversion are friendly and have a proclivity to gain power in social situations. Being in the company of others helps them to feel empowered and happy. People with a low level of extraversion are more likely to be timid and have less ability to consume in social contexts [6].
- Agreeableness: Characteristics like confidence, altruism, humanism, love, and other prosocial behaviours are included in this personality component [5]. People who score high in agreeableness are more helpful, whereas those who score low are more competitive and, at times, cunning [6].
- Neuroticism: Sadness, moodiness, and emotional unpredictability are all symptoms of neuroticism [5]. People with high levels of this component are more likely to experience mood swings, concern, irritability, and disappointment. People who score low on this characteristic are more emotionally balanced [6].

B. *Classification Algorithms*

As mentioned in section I, there are 38 categories and the collected posts and comments are classified on them, so in this study the several classification algorithms are used to classify new entries into the predefined categories or classes.

The selected classification algorithms are:

- Linear Support Vector Classification (LSVC)
- Logistic Regression (LR)
- The multinomial Naive Bayes classifier (MNB)
- Random Forest Classifier (RFC)

The accuracy is calculated for each algorithm to find the highest accuracy algorithm; the results exist in Section V.

III. Related Work

Many researchers have advocated extensive customization of user personality traits and other social traits in recommender systems, and in this section, a literature survey on personality traits is presented. Personality traits are used in many areas to improve recommender systems, and many researchers have advocated extensive customization of user personality traits and other social traits in recommender systems.

Gozde et al [7], collect data from 99,217 participants from 41 countries as part of the COVIDiSTRESS worldwide survey in order to better understand the numerous relationships between COVID-19 psychological outcomes and the "big-five" personality traits. Multigroup confirmatory factor analysis and a multilevel regression model were used to examine the data. The findings revealed that throughout the pandemic, the big-five personality characteristics were strongly linked to feelings of stress and loneliness, while some of the links were weak. Their research aids in the identification of susceptible individuals and the optimization of psychological therapy during and after the COVID-19 pandemic, where neuroticism is a role in stress and loneliness vulnerabilities, particularly during disasters.

Jose et al in [8] try give a complete knowledge of the primary user privacy problems that influence DDI. The methodology used in this study is divided into three phases: (i) a systematic literature review (SLR); (ii) in-depth interviews on user privacy concerns framed through the lenses of UGD and DDI; and (iii) topic modelling using a Latent Dirichlet allocation (LDA) model to extract insights related to the object of study. They identify 14 areas linked to the research of DDI and UGD techniques based on the findings. In addition, 14 future research topics and 7 research propositions for the study of UGD, DDI, and user privacy in digital marketplaces are offered and it finished with a discussion of the importance of user privacy in DDI in digital markets. And therefore, raised the importance of understanding the personality of users, forecasting user behaviour, and evaluating their actions.

Kamal et al in [9] offered a new deep learning-based personality prediction and classification model that combines data and classifier-level fusion. Using prominent pre-trained language models such as Elmo, ULMFiT, and BERT, this approach gained from transferring learning to natural language processing. The suggested model indicates the effectiveness of the strategy as a potential personality prediction model. Their proposed model proved to be more accurate than previous ones with 1.25 per cent and 3.12 per cent, respectively.

Valanarasu in [10] proposed a new machine learning method for predicting people's personalities based on the digital footprint of social media. During COVID-19, the suggested model may be verified for each job seeker by registering online for each organization. The suggested algorithm uses dynamic multi-context information, such as account information from Facebook, Twitter, and YouTube, among other sites. Personality prediction was more accurate than other available approaches. Even though human reasoning changes seasonally, the suggested algorithm regularly outperforms other current traditional techniques to predicting human cognition.

Masud [11] surveyed a chosen number of Facebook users based on the Short Dark Triad (SD3) model. Random Forest, Support Vector Machine, and Nave Bayes algorithms were used to categorize dark triads like Machiavellianism, Self-Centricity, and Psychopathy in this study. Compared to Random Forest and SVM, Nave Bayes produced superior results. However, showing psychopathy and self-centeredness is more difficult to detect than Machiavellianism, according to their research.

Yang Li et al in [12] proposed a new multi-task learning framework that predicted personality traits and emotional behaviours at the same time, based on the well-known correlation between personality traits and emotional attitude. It also empirically evaluated and described various information exchange mechanisms between the two tasks.

Fatemeh et al in [13] proposed a deep learning-based method for the detecting personalities from text. Convolutional neural networks (CNNs) were utilized, which have previously been found to be effective in natural language processing and personality recognition. A series of tests were used to validate their suggested technique using an essay dataset, and the empirical findings indicated that it outperformed machine learning and deep learning methods for personality detection tasks.

Ghina in [14] proposed a model to predict the personalities of Twitter users with better performance than other prediction systems. It was done by an online poll utilizing the Big-five Inventory (BFI) survey, which was sent to 295 Twitter users and collected 511,617 tweets. The suggested model employed a Support Vector Machine (SVM) to test two alternative semantic approaches that combined SVM and BERT. The results reveal that combining these two approaches yields an accuracy score of 79.35 percent, with LIWC deployment improving the accuracy score by up to 80.07 percent.

Jianshan et al in [15] proposed a new way to formalize personality as a digital twin model by observing content postings and user preference behaviours. The deep neural multitask learning network model uses two forms of data representation to estimate a user's personality. Experiments suggest that integrating the two forms of data can increase personality prediction accuracy.

Junaid et al in [16] proposed a model to classify the input text into psychopathic and nonpsychopathic features. The majority of prior work on psychopath identification has been done in the realm of psychology, employing conventional methodologies like the SRP III method and small dataset sizes. As a result, they are motivated to create advanced computer models in the field of text analysis to detect psychopaths. This study investigates sophisticated deep learning approaches, such as attention-based BILSTMs, for expanding dataset size and effectively classifying input text into psychopath and nonpsychopath classes in order to discover psychopaths.

At the conclusion of this section, it must be affirmed that understanding personality is critical to target users with appropriate advertising, detect the dark side of the personality, and assist in the recruitment process, in which some companies require understanding personality prior to employment process acceptance in order to determine whether or not a candidate has aggressive behaviour. In this project, machine learning techniques will be utilized to create a model that predicts psychological personality scores based on user-generated content utilizing the Big-five model's multiple elements and metrics. The study's hurdles include gathering data and estimating a single user's precise score in each of the big-five factors.

IV. Proposed Model

Fig. 2 shows the suggested methodology. User-generated material (posts) is sent to the preparation phase, which includes data cleaning, data transformation, and text conversion to vectors so that the algorithms can develop a prediction model, manage missing values, and remove stop words. Model of this study has two main phases Machine learning phase and personality prediction phase. The dataset of users' posts contains two main characteristics (post and post category), as shown in table I. The features are then passed on to the machine learning methods phase, which uses 80% of the dataset for training and 20% for testing.

Then move to the personality prediction phase After building the classification model, which used to predict the category of the posts for each user. The result of this phase is a new csv file with post, predicted post category, and five new columns containing discrete values for each of the big-five traits for each post category; values from 0 to 5, which 0 means that a specific big-five trait does not meet that category and 5 means that a specific big-five trait strongly meets that category, a snap shot of the csv file is shown in Fig. 3.

Finally, the model can detect the user personality between the big-five by classifying his posts and then calculating the score as shown in Table II. Table II displays an example of personality prediction results for one user posts.

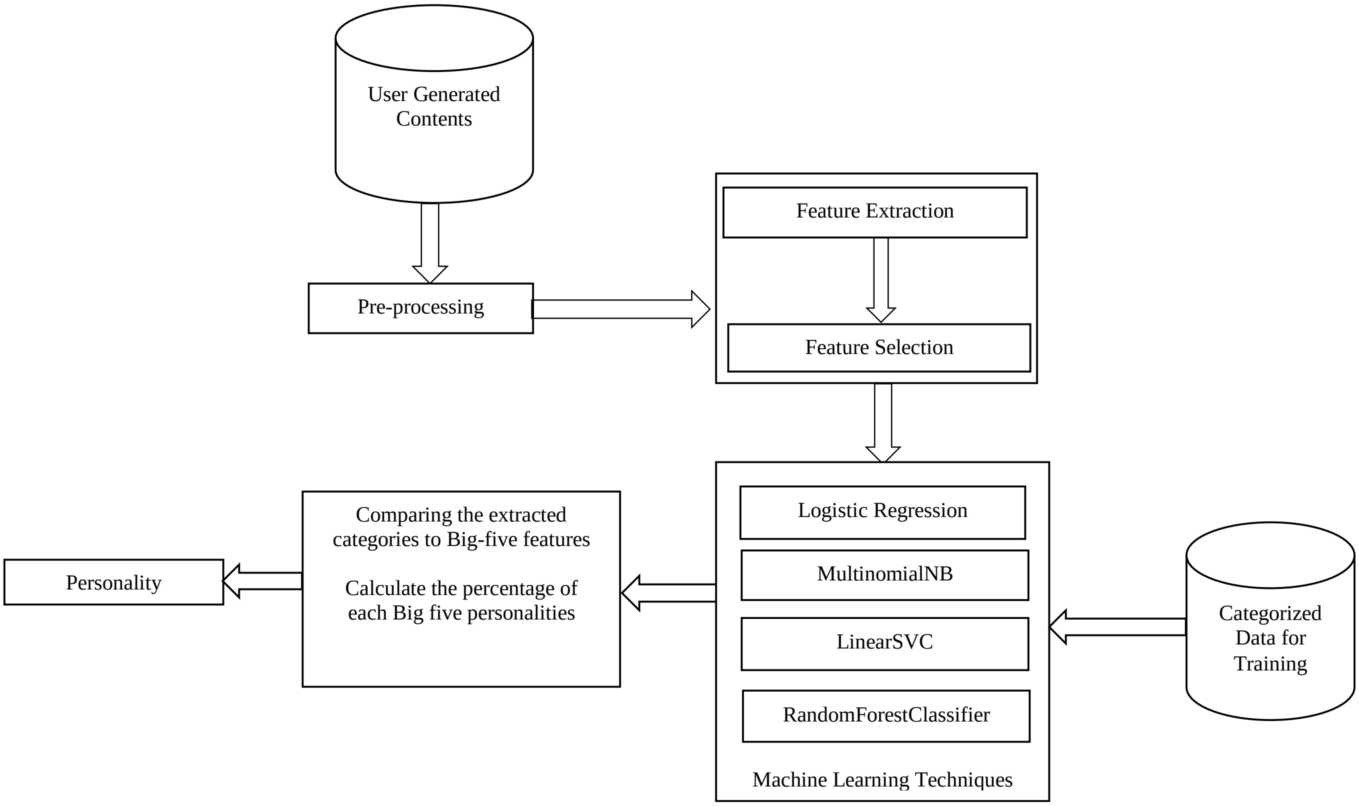
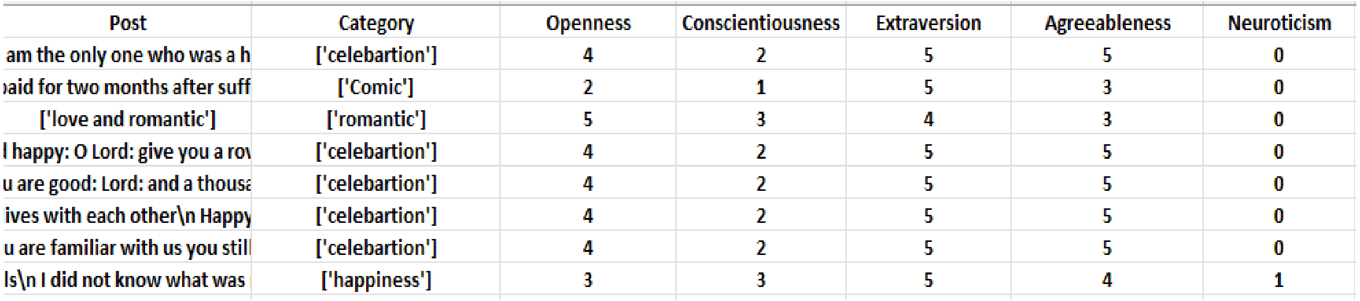

User Generated Contents
Pre-processing
Feature Extraction
Feature Selection
Logistic Regression
MultinomialNB
LinearSVC
RandomForestClassifier
Machine Learning Techniques
Categorized Data for Training
Comparing the extracted categories to Big-five features
Calculate the percentage of each Big five personalities
Personality

Fig. 2. The Proposed Model.

TABLE I. POST AND ITS CATEGORY

Post	Category
['Eman Tuhami: I swear if I said all the words of the world I cannot give you your place. During the four years college: with the feeling that I am your daughter happiness and happy day.\n Thousand million Congratulations my beloved makes us a good husband for you.']	['celebration']
['When you get paid for two months after suffering of waiting']	['Comic']



Post	Category	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
am the only one who was a h	['celebartion']	4	2	5	5	0
›aid for two months after suff	['Comic']	2	1	5	3	0
['love and romantic']	['romantic']	5	3	4	3	0
l happy: O Lord: give you a ro\	['celebartion']	4	2	5	5	0
u are good: Lord: and a thous:	['celebartion']	4	2	5	5	0
ives with each other\n Happy	['celebartion']	4	2	5	5	0
u are familiar with us you stil	['celebartion']	4	2	5	5	0
ls\n I did not know what was	['happiness']	3	3	5	4	1

Fig. 3. CSV File with Big-Five Scores.

TABLE II. EXAMPLE OF PERSONALITY PREDICTION FOR ONE USER

Big-five Trait	Score
Openess to Experience	23.859386686611817 %
Conscientiousness	20.269259536275243 %
Extraversion	23.78459237097981 %
Agreeableness	23.036649214659686 %
Neuroticism	9.050112191473449 %

TABLE III. MACHINE LEARNING ALGORITHMS ACCURACY

Model Name	Mean Accuracy
Linear SVC	89.37%
Logistic Regression	87.11%
Multinomial NB	78.81%
Random Forest Classifier	54.44 %

V. EXPERIMENTAL RESULTS

Gathered dataset comes from the user's Facebook profile, post data, and medium post titles, and is based on a merged dataset. The final dataset comprises all of the Facebook postings as well as their categories. Dataset has 38 categories that assist in obtaining more precise findings. The total number of postings in the final dataset is 27,182. The categories are verified by the psychologist Dr. Noha Gamal.

Then the model applied the machine learning algorithms LR, MNB, LSVC, and RFC for predicting the post category for each post. The results shows that LSVC achieved the highest accuracy with 89.37% while the RFC got the lowest accuracy with 54.44%, the other results are shown in Table III, Fig. 4 and Fig. 5.

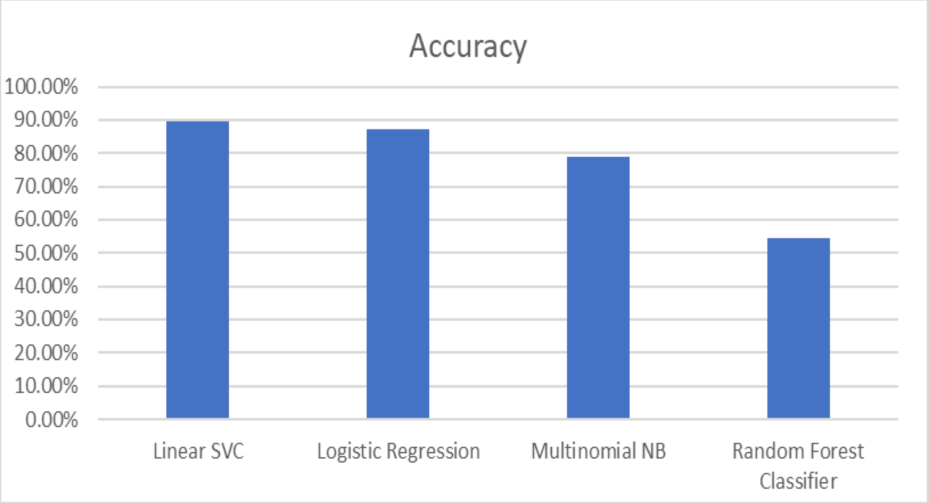


Fig. 4. Comparison of Model Performance.

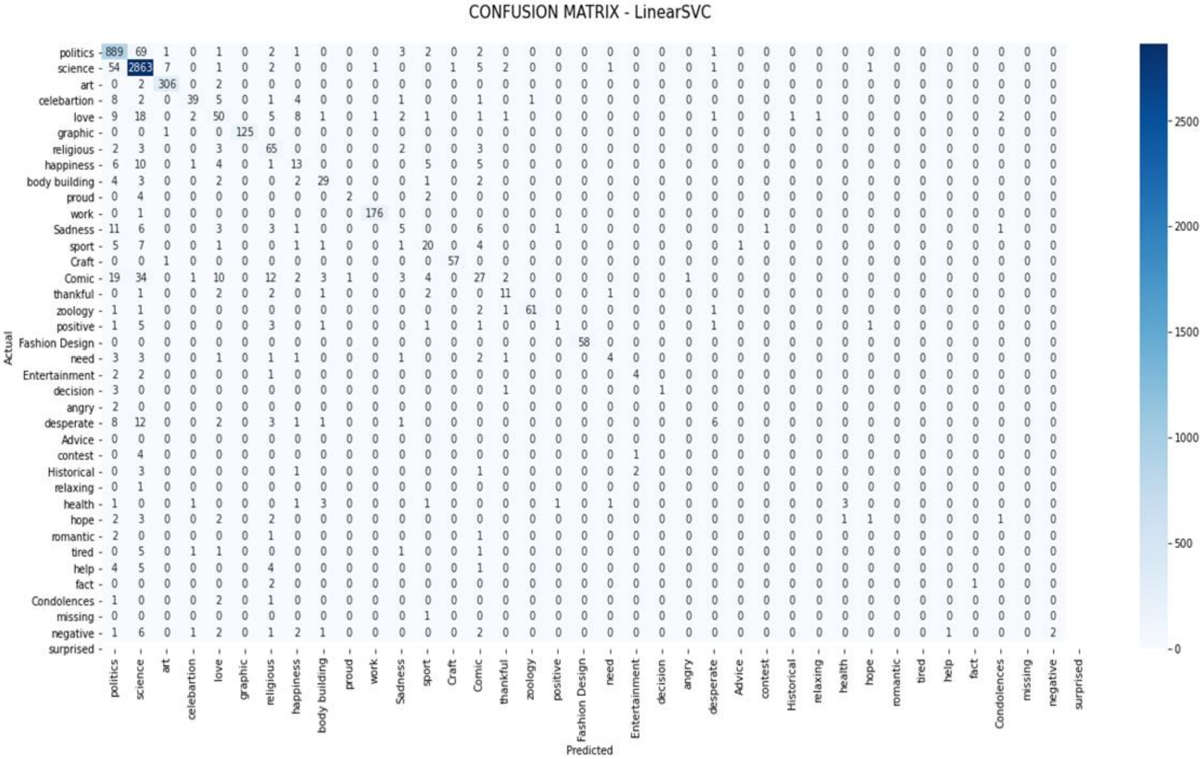


Fig. 5. LSVC Confusion Matrix.

Our achieved accuracy proves the success of the proposed model by achieving more accuracy than previous work mentioned in the previous section.

The classification results are presented and explained in this section, and Table III shows a comparison of model performance. To forecast each user's personality traits, each user's postings are sent to the model as a distinct document, and the results identify the user's personality of the big-five.

VI. Conclusion and Future Work

Understanding personality qualities is vital for appropriately treating everyone based on his or her personality traits. The proposed model can be used to target users with appropriate advertising, detect the dark side of the personality, and aid in the recruitment process, in which some companies need to understand personality before accepting the employment process to know if this person has aggressive behaviors or not. The proposed model provesc the success of identifying user’s personality according his actions and contents in social networks. The experimental results of this study represent an improvement because it can estimate the precise score of one user in each of the big-five factors. The proposed model, LinearSVC, has an accuracy of 89.37 percent. In future work, the proposed model is planned to use deep learning to improve accuracy.

References

[1] Cynthia CS Liem, Markus Langer, Andrew Demetriou, Annemarie MF Hiemstra, Achmadnoer Sukma Wicaksana, Marise Ph Born, and Cornelius J König, "Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In Explainable and interpretable models in computer vision and machine learning", Springer, 2018, pages 197–253.

[2] Jang KL, Livesley WJ, and Vernon PA, "Heritability of the big five personality dimensions and their facets: a twin study", J Pers, 1996.

[3] Han, S., Huang, H., and Tang, Y., "Knowledge of words: an interpretable approach for personality recognition from social media. Knowledge-based systems", Elsevier, 2020.

[4] Schultz, D., and Schultz, S. E, "Psychology and work today: pearson new international edition coursesmart etextbook", Routledge, 2015.

[5] RA Power and M Pluess, "Heritability estimates of the Big Five personality traits based on common genetic variants", Translational Psychiatry, 2015.

[6] Ebru Gökaliler, Ozlem Alikilic, and Fırat Tufan, "Trust and Media: Reflection of the Big Five Factor Personality Traits on COVID-19 Pandemic Communication", Turkish Review of Communication Studies, 2022.

[7] Gozde Ikizer, Marta Kowal, Ilknur Dilekler Aldemir, Alma Jeftic, Aybegum Memisoglu-Sanli, Arooj Najmussaqib, David Lacko, Kristina Eichel, Fidan Turk, Stavroula Chrona, Oli Ahmed, Jesper Rasmussen, Raisa Kumaga, Muhammad Kamal Uddin, Vicenta Reynoso-Alcantara, Daniel Pankowski, and Tao Coll-Martín, "Big Five traits predict stress and loneliness during the COVID-19 pandemic: Evidence for the role of neuroticism", Elsevier,2022.

[8] Jose Ramon Saura, Domingo Ribeiro-Soriano, and Daniel Palacios-Marques, "From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets", Elsevier, 2021.

[9] Kamal El-Demerdash, Reda A. El-Khoribi, Mahmoud A. Ismail Shoman, and Sherif Abdou, "Deep learning based fusion strategies for personality prediction", Elsevier, 2022.

[10] Mr. R. Valanarasu, "Comparative Analysis for Personality Prediction by Digital Footprints in Social Media", Journal of Information Technology and Digital World, 2021, Pages: 77-91.

[11] Shakil Mahmud, Masud Rana, Fahim Rubaiyat Zahir, and Mohammad Rezwanul Huq, "Detection of Antisocial Personality Based on Social Media Data", Springer, 2021.

[12] Yang Li, Amirmohammad Kazameini, Yash Mehta, and Erik Cambria, "Multitask Learning for Emotion and Personality Detection", IEEE, 2021.

[13] Fatemeh Mohades Deilami, Hossein Sadr, and Mohzdeh Nazari, "Using Machine Learning-Based Models for Personality Recognition", arXiv, 2022.

[14] Ghina Dwi Salsabila and Erwin Budi Setiawan, "Semantic Approach for Big Five Personality Prediction on Twitter", Rumah Jurnal Elektronik Ikatan Ahli Informatika Indonesia, 2021.

[15] Jianshan Sun, Zhiqiang Tian, Yelin Fu, Jie Geng, and Chunli Liu, "Digital twins in human understanding: a deep learning-based method to recognize personality traits", International Journal of Computer Integrated Manufacturing, 2020.

[16] Junaid Asghar, Saima Akbar, Muhammad Zubair Asghar, Bashir Ahmad, Mabrook S. Al-Rakhami, and Abdu Gumaei, "Detection and Classification of Psychopathic Personality Trait from Social Media Text Using Deep Learning Model", Hindawi, 2021.

Customer Profiling Method with Big Data based on BDT and Clustering for Sales Prediction

Kohei Arai[1], Zhan Ming Ming[2], Ikuya Fujikawa[3],Yusuke Nakagawa[4], Tatsuya Momozaki[5], Sayuri Ogawa[6]
Information Science Dept, Saga University, Saga City, Japan[1]
SIC Co., Ltd, Hakata-ku, Fukuoka City, Fukuoka, Japan[2, 3, 4, 5, 6]

***Abstract*—We propose a method for customer profiling based on Binary Decision Tree: BDT and k-means clustering with customer related big data for sales prediction; valuable customer findings as well as customer relation improvements. Through the customer related big data, not only sales prediction but also categorization of customers as well as Corporate Social Responsibility (CSR) can be done. This paper describes a method for these purposes. Examples of the analyzed data relating to the sales prediction, valuable customer findings and customer relation improvements are shown here. It is found that the proposed method allows sales prediction, valuable customer findings with some acceptable errors.**

Keywords—Customer profiling; binary decision tree: BDT; corporate social responsibility (CSR); k-means clustering; sales prediction; valuable customer findings

I. Introduction

When "analyzing customer data", it must be understood what kind of data exists, firstly. In addition, since the data in the database is used for various purposes, it is not always stored in a format suitable for analyzing customer data. Therefore, the first thing to do is to get an overall picture of customer data by profiling. Here, we propose to assume a certain data format and think based on that data format.

This is a data retention format that is generally used as target data for data mining. The retention format is a simple table with customer numbers in the vertical rows and all possible customer attributes and indicators in the horizontal columns. This table may be physically created as a database table, or it may be considered as an image when starting the analysis work just by assuming such a format.

Since the target of the analysis is the customer, first, this format is used so that the information about each customer is listed in each line. On the other hand, the attributes and index values arranged in the column are called variables. It is a "variable" meaning a "changing numerical value" because it takes a different value for each customer. The variable means the standard for looking at the customer and defines "from what point of view the customer is understood".

To perform profiling, let's look at each variable independently. Each variable is classified as either a quantitative variable or a qualitative variable. In the case of quantitative variables, understand the characteristics of variable values by using representative values such as average value, minimum value, and maximum value. Also, by understanding the distribution, it becomes possible to understand how the variable values are concentrated / distributed.

Once we understand the characteristics of each variable by profiling a single variable, the next step is to understand the relationships between the variables. A scatter-like visualization of each customer in dots makes it possible to understand whether the two variables are in direct proportion, inversely proportional, or completely random. When deciphering the relationship between variables, it is necessary to proceed with the analysis with a certain purpose, not just a visual and intuitive grasp. Data mining can be considered as one of the methods.

Large analytical datasets are vertically compressed by age group. In other words, customers are classified based on a single variable (here, age), and the tendency of each classification (= segment) is grasped. In addition to this, we have added a row for the population and a new index for the number of customers. This allows us to compare with the whole customer and understand the size of each segment. At this time, since each row to be analyzed is aggregated from customer (single customer) to segment (multiple customers), the variables to be used are average value, minimum value, maximum value, total value, composition ratio, etc. It will be converted into an index and compared. And the difference in the index value is the characteristic of the segment.

If it can be understood what kind of products it is contracting / purchasing for each segment, what kind of channel it is using, what kind of usage pattern it is using, etc. It can be understood whether it has such characteristics. With this, the profiling work is almost completed, and we have grasped what kind of variable characteristics each customer group has. The next task is analysis for "behavior". Considering that the target of analysis is the customer, the "behavior" here is an action to the customer, specifically a campaign activity.

Although there are many methods for customer profiling, performance is not good enough. Also, it is rare the method which allows prediction of sales based on the customer profiling. In this paper, we intend to improve the customer profiling performance and to propose a method for sales prediction based on Deep Learning.

There are prediction related research works as follows,

Probabilistic cellular automata-based approach for prediction of hot mudflow disaster area and volume is proposed [1]. Also, new approach of prediction of Sidoarjo hot mudflow disaster area based on probabilistic cellular automata is proposed [2]. These prediction methods are expanded to GIS based 2D cellular automata approach for prediction of forest fire spreading [3].

Cell based GIS as Cellular Automata: CA for disaster spreading prediction and required data systems is proposed [4]. The method is applied to hot mudflow prediction area model and simulation based cellular automata for LUSI and plume at Sidoarjo East Jawa [5].

Comparative study between Eigen space and real space-based image prediction methods by means of autoregressive model is conducted [6]. Also, comparative study on image prediction methods between the proposed morphing utilized method and Kalman filtering method is conducted [7].

Another prediction method for time series of imagery data in Eigen space is proposed [8] together with image prediction method with non-linear control lines derived from Kriging method with extracted feature points based on morphing [9]. Cell based GIS as cellular automata for disaster spreading predictions and required data systems are developed [10].

Prediction method of El Nino Southern Oscillation: ENSO event by means of wavelet-based data compression with appropriate support length of base function is proposed [11]. On the other hand, Question Answering: Q/A for collaborative learning with answer quality prediction is created [12].

Wildlife damage estimated and prediction using blog and tweet information is conducted in [13]. Meanwhile, prediction method for large diatom appearance with meteorological data and MODIS derived turbidity as well as chlorophyll-a in Ariake Bay area in Japan is proposed [14].

Method for thermal pain level prediction with eye motion using SVM is proposed in [15]. Meanwhile, prediction method for large diatom appearance with meteorological data and MODIS derived turbidity and chlorophyll-a in Ariake bay area in Japan [16].

Smartphone image based agricultural product quality and harvest amount prediction method is proposed [17]. On the other hand, data retrieval method based on physical meaning and its application for prediction of linear precipitation zone with remote sensing satellite data and open data is proposed and validated with the actual data [18].

Recursive Least Square: RLS method-based time series data prediction for many missing data is proposed [19]. Meanwhile, prediction of isoflavone content in beans with Sentinel-2 optical sensor data by means of regressive analysis is proposed and validated with the actual data [20].

In this connection, we propose a method for customer profiling based on Binary Decision Tree: BDT and K-means clustering with customer related big data for sales prediction and valuable customer findings as well as customer relation improvements. The scikit-learn of BDT[1] is used in the study. K-means clustering of scikit-learn[2] is also used.

The following section describes research background followed by the proposed method. Then example of the experimental actual data of sales is described as a validation of the proposed method followed by conclusion with some discussions.

II. Research Background

A. Importance of Customer Profiling

Customer profiling is important for improving sales environment and customer relation. In order to create customer profiles, customer clustering is needed, first. Then valuable customers could be found. These activities help customer need surveillance together with customer satisfaction as shown in Fig. 1.

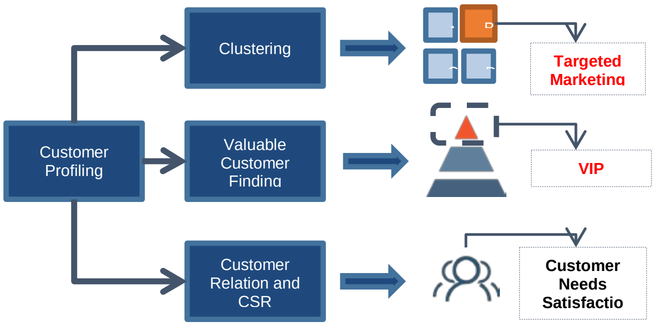

Clustering
Targeted Marketing
Customer Profiling
Valuable Customer Finding
VIP
Customer Relation and CSR
Customer Needs Satisfactio

Fig. 1. Importance of Customer Profiling.

B. Parameters for Customer Profiling

There are 7 layers of the parameters for customer profiling as shown in Table I. In the layers, typical data labels as shown in Fig. 2 must be considered.

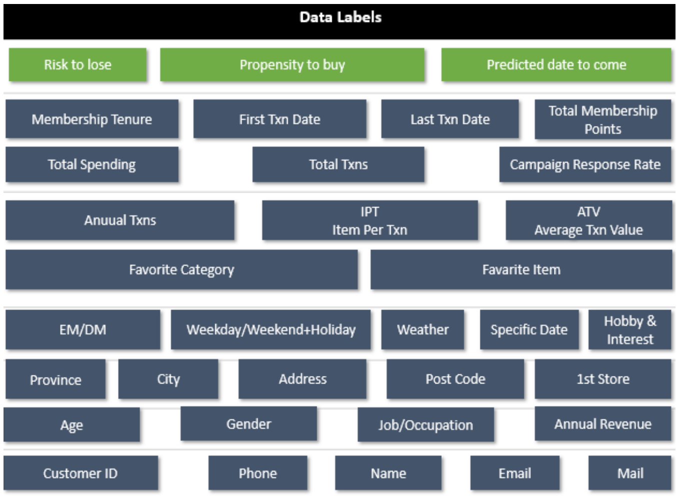
The parameters in each layer are shown in Fig. 2, the parameters in Fig. 2 shows all the possible parameters, some of them are influencing to sales prediction. The first layer includes customer information while the second layer includes demographics information. Meanwhile, the third layer includes geographic information while the fourth layer includes information on customers' preference and interest. On the other hand, the fifth layer includes shopping pattern while the sixth layer includes brand affinity. The seventh layer includes purchase action related information such as risk to lose, propensity to buy, predicted date to come.

TABLE I. Parameters for Sales Prediction

Tier 7. Prediction	7th
Tier 6. Brand Affinity	6th
Tier 5. Shopping Pattern	5th
Tier 4. Preference & Interests	4th
Tier 3. Geographic	3rd
Tier 2. Demographics	2nd
Tier 1. Recognition & Contact	1st

[1] https://scikit-learn.org/stable/modules/tree.html

[2] https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1



Data Labels

Risk to lose | Propensity to buy | Predicted date to come

Membership Tenure | First Txn Date | Last Txn Date | Total Membership Points

Total Spending | Total Txns | Campaign Response Rate

Anuual Txns | IPT Item Per Txn | ATV Average Txn Value

Favorite Category | Favarite Item

EM/DM | Weekday/Weekend+Holiday | Weather | Specific Date | Hobby & Interest

Province | City | Address | Post Code | 1st Store

Age | Gender | Job/Occupation | Annual Revenue

Customer ID | Phone | Name | Email | Mail

Fig. 2. Data Labels (Parameters) in the Assigned Layers.

III. Proposed Sales Prediction Method

A. Sales Prediction

Fig. 3 shows the proposed process flow of sales prediction. The customer profile, including the information on gender, age, address, occupation, education etc. must be made available first. Then it is followed by segmentation. After that, AI models are created, and finally total sales is going to be predicted.

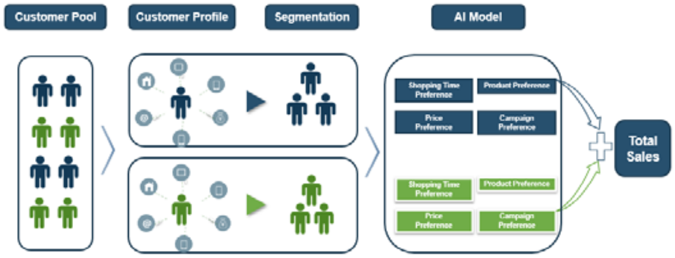


Fig. 3. Proposed Process Flow of Sales Prediction.

Fig. 4 also shows the details of sales prediction targets. The next visit date, types of sales products and sales are predicted with the AI models. Time, product, price preferences are modeled by AI through learning processes of deep learning.

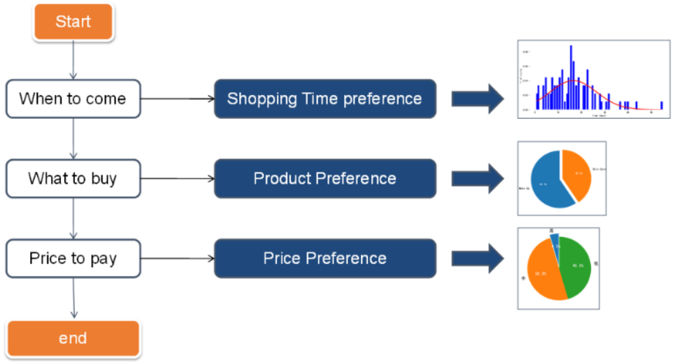
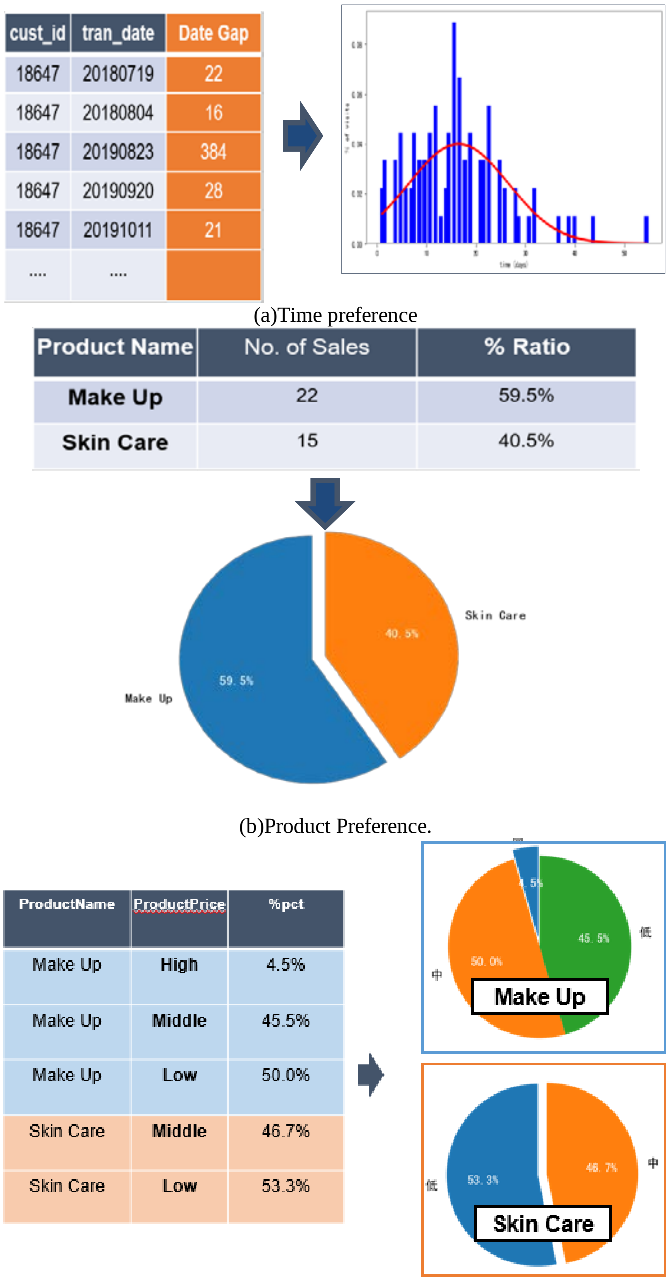


Fig. 4. Details of Sales Prediction Targets.

B. Sales Prediction with Deep Learning

Fig. 5 shows the preferences of the time, the product, and the price for the specific customer. The customer is specified through the previous processes with deep learning. All the preferences are obtained from the sales information such as shown in Table II. This is a just example of the specific customer's sales data. Total sales can also be predicted through the learned deep learning. In the proposed Deep Learning, TensorFlow Keras is used. Two hidden layers are implemented with the number of neurons of 250. Also, ReLU is used for activation function together with the fully connected layer with SoftMax function.



cust_id	tran_date	Date Gap
18647	20180719	22
18647	20180804	16
18647	20190823	384
18647	20190920	28
18647	20191011	21
....	....	

(a)Time preference

Product Name	No. of Sales	% Ratio
Make Up	22	59.5%
Skin Care	15	40.5%

(b)Product Preference.

ProductName	ProductPrice	%pct
Make Up	High	4.5%
Make Up	Middle	45.5%
Make Up	Low	50.0%
Skin Care	Middle	46.7%
Skin Care	Low	53.3%

(c)Price Preference.

Fig. 5. Preferences of the Time, the Product and the Price.

IV. Experimental Results from Sales Prediction

A. Preparation of Data for Deep Learning

The first thing we must do is to check the missing data in the customer information. Fig. 6 shows the percentage ratio of the missing data by item by item. The customer rank is the most frequently missing data item followed by e-mail addresses of the customer. Therefore, we must conduct deep learning processes without the missing data. The second thing we must do is to standardization of sales data. The ranges are different from each other so that the sales data are standardized with mean and standard deviation so as to adjust the mean and standard deviation are 0 and 1 after the standardization.

Fig. 7 shows the percentage ratio of the missing data in the sales data. As shown in Fig. 7, the highest percentage ratio of the missing data about sales is motivation to visit the shop followed by the staff ID who made services to the customer. These missing data of customer related information and the sales data must be care about in the learning processes.

TABLE II. Sales Data for the Specific Customer

Customer ID	Product ID	Product Price	No. of Sales
133034	32	Skin Care	1
133034	48	Skin Care	1
133034	80	Skin Care	1
133034	25	Make Up	1
133034	33	Make Up	1
133034	42	Make Up	1
133034	79	Make Up	2
133034	82	Make Up	3
133034	1	Make Up	2
133034	4	Make Up	2
133034	7	Make Up	1
133034	12	Skin Care	1
133034	19	Make Up	1
133034	24	Skin Care	1
133034	29	Make Up	1
133034	36	Make Up	1
133034	37	Skin Care	1
133034	40	Make Up	2
133034	52	Skin Care	1
133034	54	Make Up	1

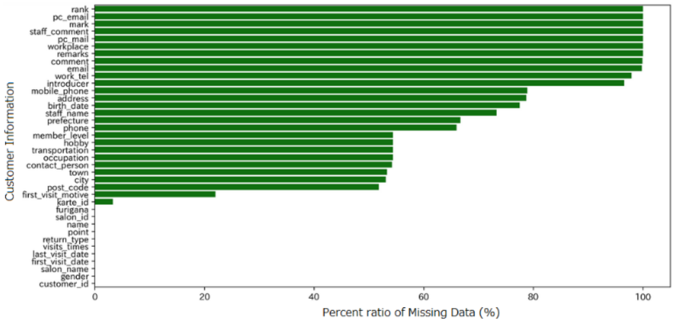


Fig. 6. Percentage Ratio of the Missing Data in the Sales Data.

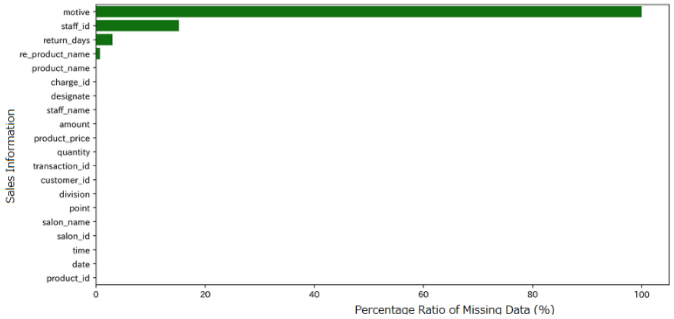


Fig. 7. Percentage Ratio of the Missing Data in the Sales Data.

Fig. 8 shows rating results of the percentage ratio of the sales type (Product name). The highest sales type is the adult hair cut (sales type No.1) followed by the gray hair dyeing (sales type No.2).

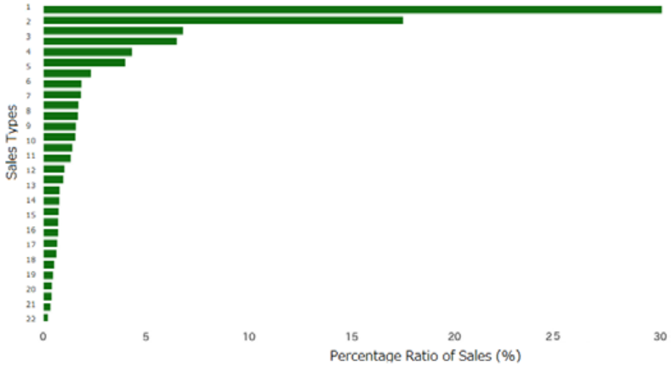


Fig. 8. Rating Results of the Percentage Ratio of the Sales Type (Product Name).

Monthly sales can also be calculated from the sales data. Fig. 9 shows the monthly sales for the specific hair salon for the period from September of 2010 to December of 2021.

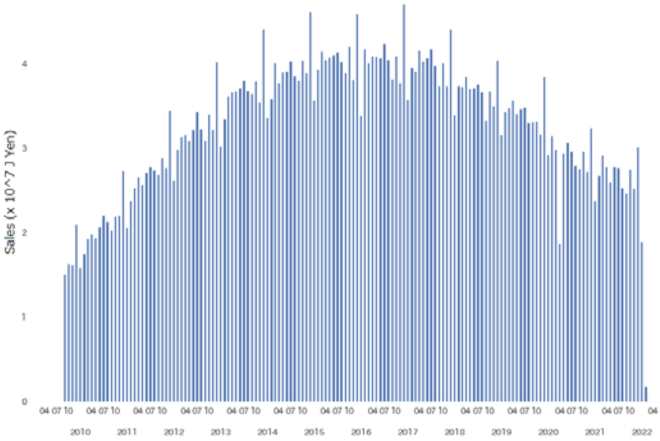


Fig. 9. Monthly Sales for the Specific Hair Salon.

Fig. 10 shows the revisit day interval (repeat cycle) for the specific hair salon.

As shown in Fig. 10, highest frequency ranges from 30 to 60 days. Therefore, it is found that the customer visits the specific hair salon every month to two months. Also, histogram of customers' age distribution is shown in Fig. 11. There are two peaks at around 15 to 20 years old for both male and female customers and 45 to 85 years old for female customers. This is the second layer of the customer profile, demographic information. The other layer data are also calculated from the sales data.

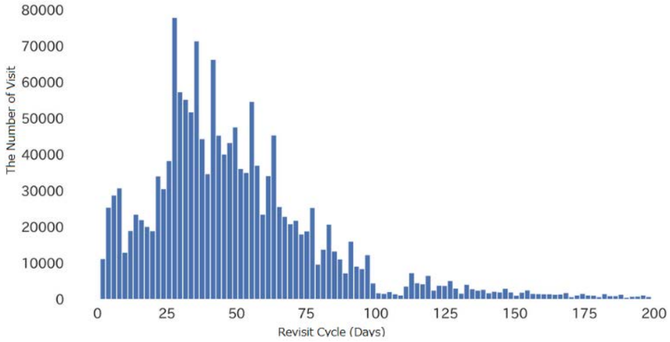


Fig. 10. Revisit Day Interval (Repeat Cycle) for the Specific Hair Salon.

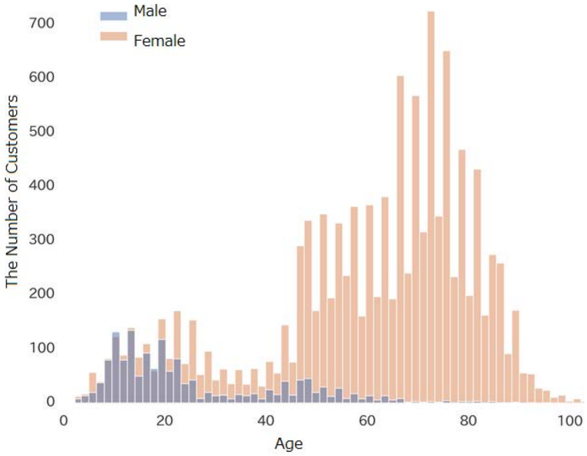


Fig. 11. Histogram of Customers' Age Distribution.

B. Customer Clustering

The number of visits to the specific hair salon is investigated as one of the parameters (feature vector) of the customer clustering.

Fig. 12 shows relation between the number of visit and the number of customers (green bars) as well as percentage ratio of revisit customers (red solid line). More than 95% of customers are repeat customers as shown in Fig. 12.

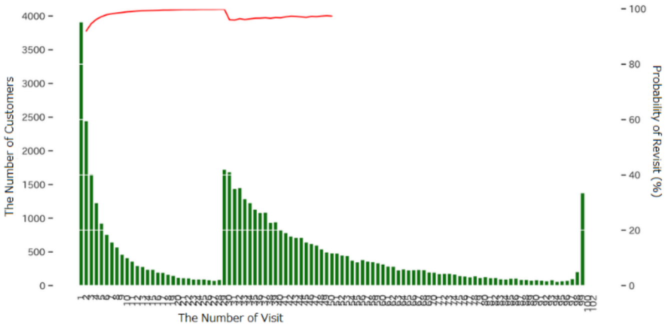
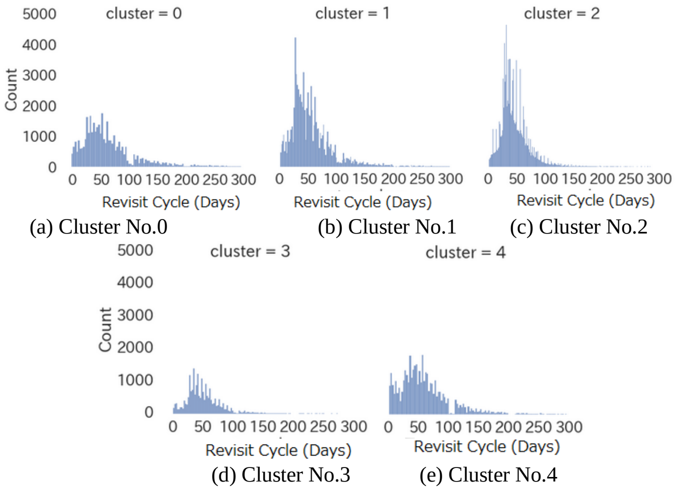


Fig. 12. Relation between the Number of Visit and the Number of Customers as well as Percentage Ratio of Revisit Customers.



(a) Cluster No.0 (b) Cluster No.1 (c) Cluster No.2

(d) Cluster No.3 (e) Cluster No.4

Fig. 13. Histogram of the Revisit Cycle (Days).

Then K-mean clustering is applied to the sales data with the number of clusters is five. Histogram of the revisit cycle (days) is shown in Fig. 13. This is one of the results from the clustering. Also, the learning processes are done with deep learning for each cluster. The cluster No. 0 is resembling to the cluster No. 4.

TABLE III. NUMBER OF CUSTOMERS FOR EACH CLUSTER

Cluster No.	The_Number_of_Customers
1	1833
4	1439
2	1360
0	1316
3	564

There is the peak of the histogram at the revisit cycle around 50. On the other hand, the cluster No. 1 is like the cluster No. 2 the peaks of the histograms of the cluster No. 1 and 2 are much higher than that of the cluster No.0 and 4 as well as the cluster No.3. The number of customers for each cluster is shown in Table III.

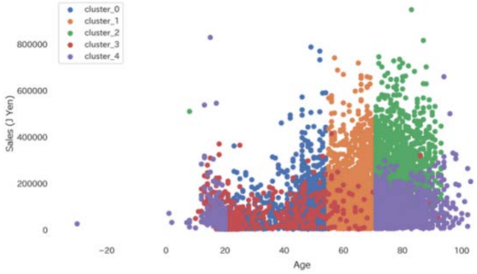
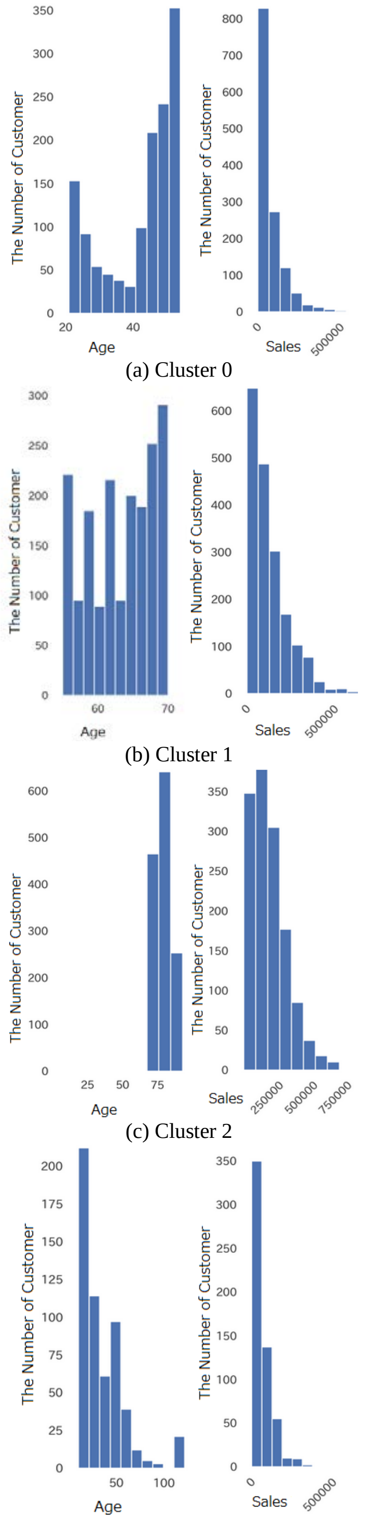


Fig. 14. Relation between the Sales Amount and Age.

C. Clustering Results

One of the clustering results is shown in Fig. 14 (relation between the age and the sales amount) for each cluster. Also, Fig. 15 shows the relations between the number of customer and age as well as sales for each cluster. As shown in Fig. 14 and 15, all the customers are divided into five clusters clearly and these clusters are well characterized with their profiles.

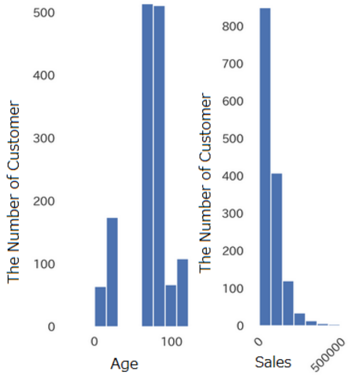


(a) Cluster 0

(b) Cluster 1

(c) Cluster 2

(d) Cluster 3



(e) Cluster 4

Fig. 15. Relations between the Number of Customer and Age as well as Sales.

Fig. 16 shows the sales amount for each cluster. The sales amount of the cluster No. 2 is highest followed by cluster No. 1. Also, there is the big dip at the late of 2019 since the number of customers is getting down by the influence due to the COVID-19.

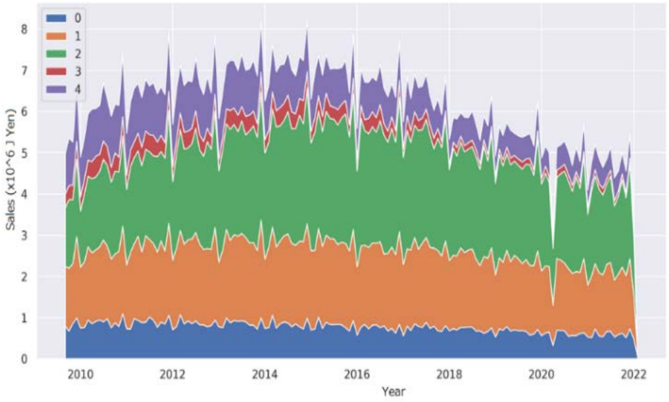


Fig. 16. Shows the Sales Amount for each Cluster.

D. Deep Learning

After the clustering, all the customers are divided into five clusters. Then the customers are segmented. Also, the number of transactions (visit), the sales amount is investigated. TensorFlow of deep learning is used for sales prediction. The training sample data is the sales data of the year from 2010 to 2013 and the validation data is the sales data of the year of 2014. The four years learning processes are conducted and then validation is done for the year of 2014. The validation result is shown in Fig. 17.

As shown in Fig. 17, prediction accuracy is not good enough (Root Mean Square Error: RMSE= 17.04: 21.25%) since the training sales data is not enough for deep learning. Also, customers’ behavior is changed by year so that the different customers’ behavior between 2010 to 2014 and 2014 induces such prediction error.

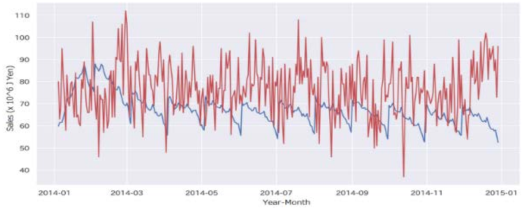


Fig. 17. Sales Prediction Result for the Year of 2014 with the Learning Processes during 2010 and 2013.

V. Conclusion

We proposed a method for customer profiling based on Binary Decision Tree: BDT and K-means clustering with customer related big data for sales prediction and valuable customer findings as well as customer relation improvements. Through the customer related big data, not only sales prediction but also categorization of customers as well as Corporate Social Responsibility: CSR can be done.

This paper describes a method for these purposes. Examples of the analyzed data relating to the sales prediction, valuable customer findings and customer relation improvements are shown in this paper. It is found that the proposed method allows sales prediction, valuable customer findings with some acceptable errors (21.25% of RMSE).

VI. Future Reaesrch Work

Further investigations are required for improvement of prediction accuracy.

Acknowledgment

The authors would like to thank to Professor Dr. Hiroshi Okumura and Professor Dr. Osamu Fukuda for their valuable discussions.

References

[1] Achmad Basuki, Tri Harsono and Kohei Arai, Probabilistic cellular automata based approach for prediction of hot mudflow disaster area and volume, Journal of EMITTER1, 1, 11-20, 2010.

[2] Kohei Arai, Achmad Basuki, New Approach of Prediction of Sidoarjo Hot Mudflow Disaster Area Based on Probabilistic Cellular Automata, Geoinformatica - An International Journal (GIIJ), 1, 1, 1-11, 2011.

[3] Kohei Arai, Achmad Basuki, GIS based 2D cellular automata approach for prediction of forest fire spreading, International Journal of Research and Reviews on Computer Science, 2, 6, 1305-1312, 2011.

[4] Kohei Arai, Cell based GIS as Cellular Automata for disaster spreading prediction and required data systems, CODATA Data Science Journal, 137-141, 2012.

[5] Kohei Arai, A.Basuki, T.Harsono, Hot mudflow prediction area model and simulation based cellular automata for LUSI and plume at Sidoarjo East Jawa, Journal of Computational Science (Elsevior) 3,3, 150-158, 2012.

[6] Kohei Arai, Comparative Study between Eigen Space and Real Space Based Image Prediction Methods by Means of Autoregressive Model, International Journal of Research and Reviews in Computer Science (IJRRCS) Vol. 3, No. 6, 1869-1874, December 2012, ISSN: 2079-2557.

[7] Kohei Arai, Comparative Study on Image Prediction Methods between the Proposed Morphing Utilized Method and Kalman Filtering Method, International Journal of Research and Reviews in Computer Science (IJRRCS) Vol. 3, No. 6, 1875-1880, December 2012, ISSN: 2079-2557.

[8] Kohei Arai Prediction method for time series of imagery data in eigen space, International Journal of Advanced Research in Artificial Intelligence, 2, 1, 12-19, (2013).

[9] Kohei Arai Image prediction method with non-linear control lines derived from Kriging method with extracted feature points based on morphing, International Journal of Advanced Research in Artificial Intelligence, 2, 1, 20-24, (2013).

[10] Kohei Arai, Cell based GIS as cellular automata for disaster spreading predictions and required data systems, Advanced Publication, Data Science Journal, Vol.12, WDS 154-158, 2013.

[11] Kohei Arai, Prediction method of El Nino Southern Oscillation event by means of wavelet based data compression with appropriate support length of base function, International Journal of Advanced Research in Artificial Intelligence, 2, 8, 16-20, 2013.

[12] Kohei Arai, Anik Nur Handayani, Question Answering for collaborative learning with answer quality prediction, International Journal of Modern Education and Computer Science, 5, 5, 12-17, 2013.

[13] Kohei Arai, Shohei Fujise, Wildlife Damage Estimated and Prediction Using Blog and Tweet Information, International Journal of Advanced Research on Artificial Intelligence, 5, 4, 15-21, 2016.

[14] Kohei Arai, Prediction method for large diatom appearance with meteorological data and MODIS derived turbidity as well as chlorophyll-a in Ariake Bay area in Japan, International Journal of Advanced Computer Science and Applications IJACSA, 8, 3, 39-44, 2017.

[15] Kohei Arai, Method for Thermal Pain Level Prediction with Eye Motion using SVM, International Journal of Advanced Computer Science and Applications IJACSA, 9, 4, 170-175, 2018.

[16] Kohei Arai, Prediction method for large diatom appearance with meteorological data and MODIS derived turbidity and chlorophyll-a in Ariake bay area in Japan, International Journal of Advanced Computer Science and Applications IJACSA, 10, 9, 39-44, 2019.

[17] Kohei Arai, Osamu Shigetomi, Yuko Miura, Satoshi Yatsuda, Smartphone image based agricultural product quality and harvest amount prediction method, International Journal of Advanced Computer Science and Applications IJACSA, 10, 9, 24-29, 2019.

[18] Kohei Arai, Data Retrieval Method based on Physical Meaning and its Application for Prediction of Linear Precipitation Zone with Remote Sensing Satellite Data and Open Data, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 10, 56-65, 2020.

[19] Kohei Arai, Kaname Seto, Recursive Least Square: RLS Method-Based Time Series Data Prediction for Many Missing Data, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 11, 66-72, 2020.

[20] Kohei Arai, Prediction of Isoflavone Content in beans with Sentinel-2 Optical Sensor Data by Means of Regressive Analysis, Proceedings of SAI Intelligent Systems Conference, IntelliSys 2021: Intelligent Systems and Applications pp 856-865, 2021.

Authors' Profile

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 since 2008 then he is now award committee member of ICSU/COSPAR. He wrote 77 books and published 670 journal papers as well as 500 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

A New Hybrid-Heuristic Approach for Vertex *p*-Median Location Problems

Hassan Mohamed Rabie[1]
Institute of National Planning
Egypt

Said Salhi[2]
Kent Business School, University of Kent
Canterbury CT2 7FS

***Abstract*—In this paper, a new hybridization of a Myopic and Neighborhood approaches is proposed to solve large-size vertex *p*-median location problems. The effectiveness and efficiency of our approach are demonstrated empirically through an intensive computational experiment on large-size instances taken from TSPLib and BIRCH datasets, with the number of nodes varying from 734 to 9,976 for the former and from 9,600 to 20,000 nodes for the latter. The results show that the new approach, though relatively simple, yields better solutions compared to the ones in the literature. This demonstrates that a simpler approach that takes into account the advantages of other methods can lead to promising outcome and has the potential of being adopted in other combinatorial optimization problems.**

Keywords*—*P-median; discrete location problems; myopic heuristic; neighborhood heuristic

I. Introduction

The *p*-median location problem is one of the oldest discrete location problems. The objective of the vertex *p*-median location problem is to find the location of *p* median facilities among *n* demand points to minimize the sum of the distances between customers and their nearest median facilities. This problem is also known as the minisum vertex location problem [1]. In the uncapacitated type, each median facility is not restricted by the number of demand points/customers to serve, however, in the capacitated *p*-median location problem each median facility has a fixed capacity. In this paper, we are interested in addressing the former for the case of large-scale instances where exact methods may not be suitable. This type of location problems has many applications such as, locating the locations of the ambulances, schools, firefighters and hospitals among others [2-5].

The problem was first introduced in [6, 7], and has been proved to be an NP-hard optimization problem [8]. For large-size location problems, optimal solutions may not be reached, therefore, heuristic and metaheuristic approaches are usually the best way forward for solving these vertex *p*-median location problems [9, 10]. For more information on the vertex *p*-median location problems, see [11-13]. The vertex *p*-median location problem was formulated by ReVelle and Swain in [14] and its implementation was enhanced by Rosing et al. in [15]. The following notation is used.

Let I be the set of nodes (demand points), J the set of potential sites, and C_{ij} the distance between site i ($i \in I$) and demand point j ($j \in J$).

Let p the number of median facilities to be located and let y_i and x_{ij} the following decision variables with:

$$y_i = \begin{cases} 1, if\ a\ facility\ is\ located\ at\ candidate\ site\ i \\ \qquad\qquad 0, otherwise \end{cases}$$

and x_{ij} the fraction of the demand of customer j that is supplied from facility i.

The p-median location problem can be formulated as follows:

$$minimize\ \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \tag{1}$$

$$subject\ to\ \sum_{i \in I} x_{ij} = 1\ \forall\ j\ \in J \tag{2}$$

$$\sum_{i \in I} y_i = p \tag{3}$$

$$x_{ij} - y_i \leq 0\ \forall\ i\ \in I\ ;\ j\ \in J \tag{4}$$

$$y_i\ \in \{0\ , 1\}\ ;\ \forall\ i\ \in I \tag{5}$$

$$x_{ij}\ \geq\ 0\ ;\ j\ \in J \tag{6}$$

Equation (1) is the objective function which minimizes the sum of the total distances, (2) states that all demand at demand site (*j*) must be satisfied, (3) guarantees that exactly *p* median facilities are to be located. Equation (4) ensures that demand nodes can be only assigned to the open median facilities, (5) specifies that the location variables have to be binary, and finally, (6) requires that the assignment variables have to be non-negative.

Though there exist heuristic approaches dedicated for the vertex *p*-median location, there is no recent research that has concentrated on comparing the performance of the heuristic approaches with the state-of-the-art techniques. In this paper, we aim to fill that gap.

A. Contribution and Organization of the Paper

Here, the basic Myopic construction and Neighborhood improvement approaches are first outlined. A new Hybrid-heuristic that integrates the two above techniques is also proposed. In other words, the contribution of this study is three folds:

- Introduce a new Hybrid-heuristic approach which integrates Myopic and Neighborhood for solving large-size vertex *p*-median problems,

- Compare the performance of heuristic approaches versus the performance of recent existing approaches, and
- Generate new best solutions for large instances that can be used for benchmarking purposes or even as bounds for exact methods if need be.

The paper is organized as follows. The next Section presents a brief review of the related literature. The third Section describes some vertex *p*-median heuristic approaches, with an emphasis on the new proposed approach. In Section 4, an illustrative example is presented followed by the computational results in Section 5. The final Section outlines our conclusion and highlights some research avenues.

II. Literature Review

The difficulty of solving large-size vertex *p*-median location problems led several authors to investigate alternative approaches and techniques [16-18]. This section provides a brief overview of some work on recent large-size vertex *p*-median location problems. See [19, 20] for a review on this topic.

In [16], Avella et al. presented a Branch-and-Cut-and-Price approach, which yields reasonable solutions for instances with demand points/customers (*n*) less than 3,795. They applied their approach on the OR library instances, namely, the TSP-Lib instances [21]: Fl1400, PCB3038 and Rl5934, and the Optimal Diversity Management (ODM) instances BN1284, BN3373 and BN5535. In [22], Hansen et al. used a Variable Neighborhood Search (VNS) to find a solution for the clustering problem as a large-size vertex *p*-median location problem. They reported the results for two large-size datasets, where each dataset consists of *p* groups (clusters) of two dimensional data points generated in a square. The datasets consist of BRICH I instances and BRICH III instances [23]. The size of each dataset ranged from 10,000 demand points to 89,600 demand points. The TSP-Lib library is used to choose the second set of instances from [21]: PCB3038, RL5915, RL5934, RL11849, USA13509, IT16862, SW24978, and BM33708, where the number attached to each instance title indicates the number of demand points (*n*) in that instance. In [24], Avella et al. introduced a new Lagrangean relaxation heuristic to solve large-size *p*-median location instances. The algorithm consists of three main components: (1) Subgradient column generation, (2) Core heuristic, and (3) Aggregation procedure. The authors reported the solution of instances from BRICH (two of type BRICH I and two of type BRICH III), and also two instances from the TSP-Lib library [21], namely PCB3038 and USA13509. In [25], Irawan et al. introduced a multiphase approach that includes three parts; (1) demand points aggregation, (2) Variable Neighbourhood Search and (3) an exact approach to solve large-size unconditional and conditional *p*-median location vertex problems. The approach consists of four phases (stages). The first phase solves several aggregated problems using a ‘‘Local Search with Shaking’’ procedure to generate candidate solutions which are then used to solve a reduced location problem in Phase two using Variable Neighbourhood Search or an exact method. The new candidate solution set is then introduced as an input for the iterative learning process to tackle the aggregated *p*-median location problem in Phase three. Finally, Phase four is a post optimization phase applied to solve the original problem using a local search, starting from the best solution obtained in the previous phase. This multiphase approach is tested on three well-known datasets. The first is the BIRCH datasets (BIRCH I and BIRCH III), the second is based on the TSP-Lib library [21] which includes IT16862, SW24978, BU33708 and CH71009 whereas the third is the Circle dataset, which is a newly geometrically generated by the authors to guarantee optimal solutions and hence provides a strong comparison with the heuristic produced. In [26], Irawan and Salhi further designed a hybrid technique based on clustering and VNS with the aim to find a solution to large-size *p*-median location problems. The new approach is a multi-step methodology in which learning from previous steps is taken into account when tackling the next step. Each step consists from sub-problems which are solved by a fast procedure to produce good feasible solutions. Within each step, the solutions are grouped together to produce a new promising subset of potential medians/facilities. This is similar in principle to data mining and heuristic concentration developed by Rosing and ReVelle [27]. The proposed approach is tested on BIRCH datasets. In [3], Janáček and Kvet studied the public service system design which is formulated as a *p*-median location problem, through focusing on the approximate radial approach using dividing points. The approximate approach can be implemented using any commercial integer programming solver. The proposed approach is tested on TSP-Lib instances; RL1304, FL1400, U1432, V1748, D2103, and PCB3038. In [28], Vasilyev and Ushakov proposed a new modified hybrid sequential Lagrangian heuristic that uses a shared memory parallel implementation which can be used in suitable technology. They integrate their Lagrangian relaxation approach with a sub gradient column generation and a core selection method in combination with a simulated annealing to identify the sequences of lower and upper bounds for the optimal value. The proposed approach is also tested on BIRCH datasets. In [29], Vasilyev et al. addressed a general fault-tolerant version of the p-median location problem. The authors adapted their earlier method to determine the upper and lower bounds. They tested the proposed method on large-scale problem instances taken from TSPLIB library: JA9847, USA13509, IT16862, and SW24978.

In summary, due to the importance of the vertex *p*-median location problem and its real-life applications, a considerable amount of *p*-median location approaches have been proposed such as Branch-and-cut-and-price, Variable neighborhood search, Lagrangean relaxation, among others. Though the above approaches are promising, they are not easily and widely applicable. We therefore believe there could still be the need to propose a simple but powerful and effective heuristic-based approach.

III. Heuristics Approaches for the *p*-Median Problem

This section outlined two classes of heuristic approaches: the Myopic approach, and the Neighborhood Search approach. The Myopic approach is a construction method that builds a good solution from scratch, while the neighborhood search approach is an improvement algorithm. More details can be found in [19, 20, 30]. This will then be followed by the new hybrid myopic-neighborhood which we propose.

A. *Myopic Approach*

Myopic is the simplest greedy add (i.e., construction) heuristic approach. The Myopic method starts with an empty set of medians (vertices/points) and successively adds the candidate vertex (point) that yields the best decrease in the minisum objective function value. The process continues until the solution includes *p* median facilities [19]. The Myopic Approach is simple to understand and to implement. However, it suffers from the fact that once a facility is added, it is not removed in subsequent iterations and therefore will restrict the search space. Fig. 1 shows the commonly used Myopic pseudocode. Let us define the function $Z(J,X) = \sum_{j\in J} min_{m\in X}\{c_{mj}\},$; where X is the current set of candidate solutions. The function depends on both the set of demand nodes to be considered and the candidate locations to be used.

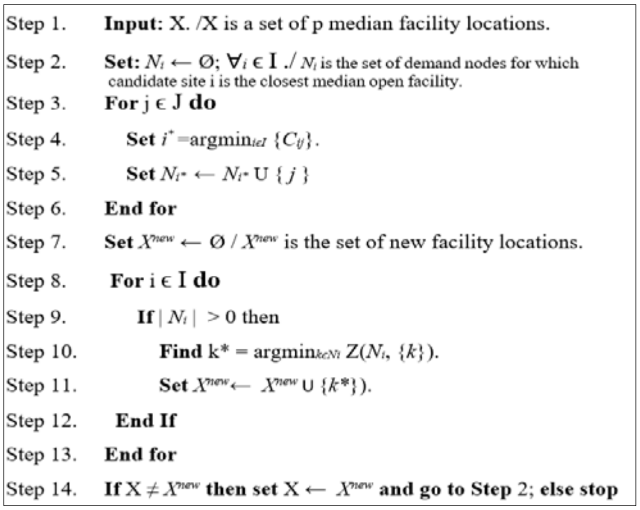
```
Step 1. Let X ← Ø. /Where X is the set of locations, starting with an empty set.
Step 2. Find $i^*$ =argmin$_{i\epsilon I}$ {Z(J, X U{i}x)}. / the best node to add to the
        solution set.
Step 3. Set X ← X U { $i$* }./ Adds that site to the solution.
.       If |X| < P, go to Step 2; else stop.
```

Fig. 1. Myopic Algorithm Pseudocode [20].

Step 1 starts by initializing the set of candidate solutions to an empty set. Step 2 relates to the best vertex (point) to be added to the candidate solution set. Step 3 adds that vertex/point to the candidate solution set. Step 4 checks if less than *p* median facilities have been added to the solution set. If so, the Myopic approach continues with Step 2; if not, the search terminates. According to Daskin [30], the solution obtained using Myopic method may not be optimal as outlined earlier. This is because once a site is chosen, it remains there which restricts the search making the solution suboptimal as the optimal solution may not necessarily have the additive property. In other words, there is no guarantee of optimality for the Myopic approach, unless we are locating only a single median facility [20].

B. *Neighborhood Approach*

Neighborhood approach attempts to improve a given solution made up of *p* candidates. It can be considered as one of the most-widely and oldest improvement mechanism [19, 30]. The approach starts with any feasible candidate solution to the vertex *p*-median location problem (For example, it could begin with the solution set identified by the Myopic approach), then the approach assigns each demand vertex to its nearest median facility. Then the one median location problem within each neighborhood is selected through examining each candidate demand point. If the solution of the one median location results in a new location for the median facility, the approach reallocates all demand points to the nearest open median facility. Otherwise (i.e., if no change for the median facility locations), the approach stops. If there is no new assignments, the approach also stops; otherwise, the search continues [20]. Fig. 2 shows the pseudocode of the Neighborhood search approach. Note that this approach was initially developed for the Weber problem (i.e, the *p*-median problem but on the plane) by [31] which is known as the locate-allocate method.



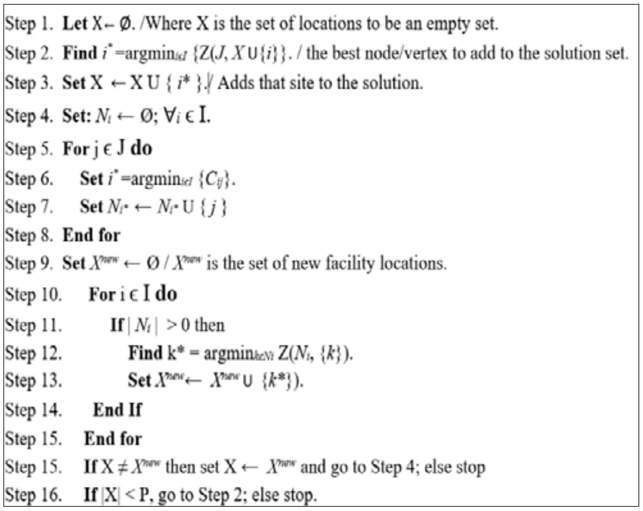
```
Step 1.   Input: X. /X is a set of p median facility locations.
Step 2.   Set: $N_i$ ← Ø; ∀$_i$ ϵ I ./ $N_i$ is the set of demand nodes for which
           candidate site i is the closest median open facility.
Step 3.   For j ϵ J do
Step 4.       Set $i^*$ =argmin$_{i\epsilon I}$ {$C_{ij}$}.
Step 5.       Set $N_{i^*}$ ← $N_{i^*}$ U { $j$ }
Step 6.   End for
Step 7.   Set $X^{new}$ ← Ø / $X^{new}$ is the set of new facility locations.
Step 8.    For i ϵ I do
Step 9.         If | $N_i$ | > 0 then
Step 10.            Find k* = argmin$_{k\epsilon Ni}$ Z($N_i$, {$k$}).
Step 11.            Set $X^{new}$← $X^{new}$ ∪ {$k$*}).
Step 12.     End If
Step 13.  End for
Step 14.  If X ≠ $X^{new}$ then set X ← $X^{new}$ and go to Step 2; else stop
```

Fig. 2. Neighborhood Search Approach Pseudocode [20].

Step 1 starts by initializing the solution with any set of *p* median facilities. Step 2 to Step 6 initialize and set the vertices/points neighborhood. Step 7 initializes a new candidate set of median facility locations. Step 8 to Step 13 the new candidate locations is found. Step 10 calculates the 1-median location problem within each neighborhood and adds that vertex/point to the solution set in Step 11.

C. *The Proposed Hybrid Myopic-Neighborhood Approach*

This section outlines the new Hybrid Myopic-Neighborhood. The main idea of the proposed approach is that, at each iteration of the Myopic and after determining the next point (vertex) to be added to the current solution; the Neighborhood approach is used as many times as possible till there is no improvement in the current candidate solution. In other words, this embedded local search acts as a filtering mechanism during the search process as the open set of facilities is augmented. Fig. 3 shows the pseudocode of the Hybrid Myopic-Neighborhood approach.



```
Step 1.  Let X← Ø. /Where X is the set of locations to be an empty set.
Step 2.  Find $i^*$ =argmin$_{i\epsilon I}$ {Z(J, X ∪{i}}. / the best node/vertex to add to the solution set.
Step 3.  Set X ← X U { $i$* }./ Adds that site to the solution.
Step 4.  Set: $N_i$ ← Ø; ∀$_i$ ϵ I.
Step 5.  For j ϵ J do
Step 6.      Set $i^*$ =argmin$_{i\epsilon I}$ {$C_{ij}$}.
Step 7.      Set $N_{i^*}$ ← $N_{i^*}$ U { $j$ }
Step 8.  End for
Step 9.  Set $X^{new}$ ← Ø / $X^{new}$ is the set of new facility locations.
Step 10.     For i ϵ I do
Step 11.         If | $N_i$ | > 0 then
Step 12.             Find k* = argmin$_{k\epsilon Ni}$ Z($N_i$, {$k$}).
Step 13.             Set $X^{new}$← $X^{new}$ ∪ {$k$*}).
Step 14.      End If
Step 15.    End for
Step 15.    If X ≠ $X^{new}$ then set X ← $X^{new}$ and go to Step 4; else stop
Step 16.    If |X| < P, go to Step 2; else stop.
```

Fig. 3. Hybrid Myopic-Neighborhood Approach Pseudocode.

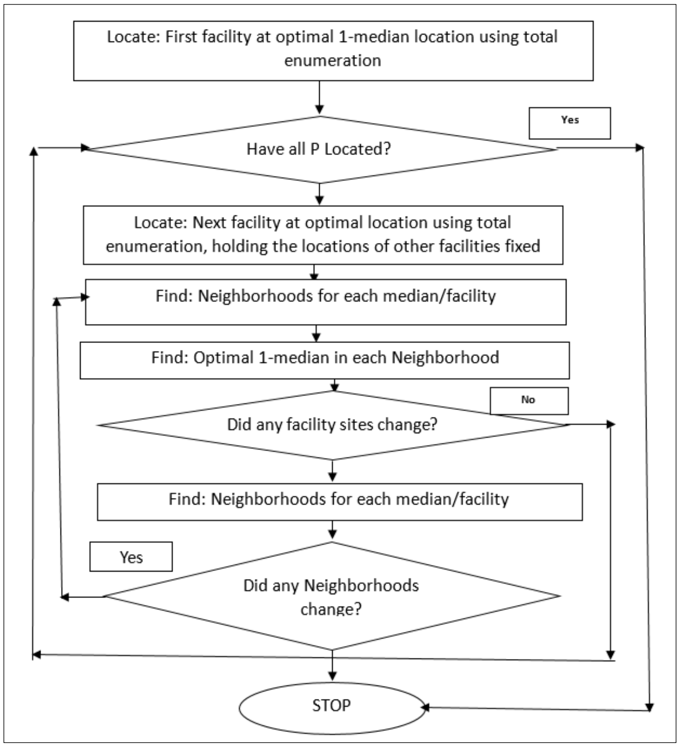

Locate: First facility at optimal 1-median location using total enumeration
Have all P Located?
Yes
Locate: Next facility at optimal location using total enumeration, holding the locations of other facilities fixed
Find: Neighborhoods for each median/facility
Find: Optimal 1-median in each Neighborhood
Did any facility sites change?
No
Find: Neighborhoods for each median/facility
Yes
Did any Neighborhoods change?
STOP

Fig. 4. A Simple Flowchart of the New Hybrid Myopic-Neighborhood Approach.

For completeness, we also provide Fig. 4 as a flowchart for the Hybrid Myopic-Neighborhood approach.

IV. Numerical Example

This section introduces a small example to illustrate the new Hybrid Myopic-Neighborhood approach to solve vertex location problem. For simplicity let us consider an example from Daskin [30] with Table I shows the distance matrix.

For the first median (*p*=1), using total enumeration, the location of the 1st median is located at vertex (point) 9.

For the second median (*p*=2), (1) applying Myopic: The location of the 2nd median is located at vertex 7 and the set of solution X= {9, 7}. According to the Myopic, we should proceed by adding a new third vertex which yields the lowest objective function value, however, the new Hybrid approach proposed applying Neighborhood approach; (2) applying Neighborhood, no change has been happened in the solution set, therefore, go to next median (i.e., *median* # 3).

For the third median (*p*=3), (1) applying Myopic, the location of the 3rd median is located at vertex 6 and the set of solution X={9,7,6}. (2) applying Neighborhood, The new location set, X^{new}={11,7,6}, a change happened at site 9 and swapped with site 11, therefore, go to step 4, and apply again Neighborhood. Now, there is no median facility changed; go to Next median, and so on. The detailed calculations can be found in Appendix A.

It is clear that using the new Hybrid approach improves the objective function value of the Myopic, which considered an added sequence approach. The best solution at *p* is not necessarily the best solution at *p*-1 with adding an additional vertex. This is because the additivity property is not satisfied. This observation was also taken advantage of when applying the ‘drop’ method instead of the ‘add’ method, as demonstrated by the flexible drop method, known as subdrop, originally developed by Salhi and Atkinson [32]. Also, it is worth noting that the new Hybrid either improves the objective function value of the Neighborhood or retain that value but never worsen it, since the Hybrid keeps starting from a better intermediate initial solution.

Table II compares the results of using the new Hybrid approach to solve the example in Table I, versus the results of Myopic, Neighborhood, Exchange, and Lagrangian approaches in [30]. The table shows that the new Hybrid Myopic-Neighborhood outperformed other heuristics, namely, Myopic, Neighborhood, and Exchange heuristics approaches. The Hybrid approach achieved optimality at all levels of medians (*q*=1,…,*p*), such as the Lagrangian exact approach. However, the new Hybrid approach is very simple to understand and relatively easier to implement in practical setting.

TABLE I. The Distance Matrix for a p-median Example from [30]

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	225	555	825	360	900	270	495	720	600	870	1005
2	150	0	220	400	380	520	330	480	420	550	610	610
3	444	264	0	216	192	360	492	336	240	696	468	468
4	990	720	324	0	612	216	1062	828	432	1116	774	612
5	120	190	80	170	0	180	125	60	120	235	185	215
6	1440	1248	720	288	864	0	1368	1008	288	1200	744	528
7	198	363	451	649	275	627	0	165	495	242	440	671
8	528	768	448	736	192	672	240	0	480	592	400	736
9	624	546	260	312	312	156	585	390	0	494	247	247
10	880	1210	1276	1364	1034	1100	484	814	836	0	418	880
11	1102	1159	741	817	703	589	760	475	361	361	0	399
12	1340	1220	780	680	860	440	1220	920	380	800	420	0

TABLE II. COMPARING THE HYBRID APPROACH VERSUS OTHER HEURISTIC AND EXACT APPROACHES

P	**Myopic**	**Exchange**	**Neighborhood**	**Lagrangian**	**Hybrid**
1	**4,772**	**4,772**	**4,772**	**4,772**	**4,772**
2	**3,145**	**3,145**	**3,145**	**3,145**	**3,145**
3	2,641	2,498	2,641	**2,438**	**2,438**
4	2,157	**1,884**	2,157	**1,884**	**1,884**
5	1,707	**1,444**	1,572	**1,444**	**1,444**
6	1,327	**1,083**	1,192	**1,083**	**1,083**
7	966	**747**	**747**	**747**	**747**
8	666	**531**	666	**531**	**531**
9	426	**366**	426	**366**	**366**
10	**210**	**210**	**210**	**210**	**210**
11	**60**	**60**	**60**	**60**	**60**
12	**0**	**0**	**0**	**0**	**0**

V. COMPUTATIONAL RESULTS

To assess the performance of the new Hybrid Myopic-Neighborhood approach to solve large-size vertex *p*-median location problems, an extensive computation experiment using two frequently types of datasets is carried out. These datasets consist of the TSP-Lib, and BIRCH instances. Computational results of the vertex *p*-median instances are listed in Table III, Table IV, Table V and Table VI. This section reports the computational results on a subset of large-size instances which have previously been used in [22, 24, 25, 28]. Computational experiments were carried out on a processor Intel(R) Core(TM) i7-8550U CPU@1.80GHz 1.99 with 8 GB of RAM, under Windows 10, 64-bit. The code was written and executed in MATLAB.

A. TSP-Lib Instances

The first dataset of instances is taken from the TSP-Lib, a travelling salesman library [21]. There are 11 instances (UY734, ZI929, MU1979, CA4663, TZ6117, EG7146, YM7663, EI8246, JA9847, GR9882, and KZ9976), where each instance is solved with *p* varying from 25 to 75 with an increment of 5. The number attached to each instance name indicate the number of demand points (*n*) of that instance. For example, UY734 contains the coordinates of 734 cities in the Uruguay. The instances are ranged in size from *n* = 734 to 9,976.

1) Comparison vs Neighborhood "NBHD" and Myopic: The computational results for the new Hybrid approach on the TSP-Lib dataset are presented in Table III alongside those obtained by the Neighborhood "NBHD" and Myopic approaches, with 'Bold' showing the best solutions. For clarity, the items in Table III are as follows:

a) n is the number of demand points,

b) p is the number of median facilities to be located,

c) z is the minisum objective function obtained by the three approaches,

d) Deviation(%): The percentage gap between a given solution and the best solution. It is computed as: *Deviation*(%)= 100*(Z_H-Z_{Best})/Z_{Best}), where Z_H and Z_{Best} correspond to the *Z* value obtained with heuristic '*H*' and the best *Z* value respectively. 'Bold' values in the table refer to the best solutions [25].

e) Time (*Sec*); Time in seconds. We should notice that the time of Neighborhood "NBHD" is significantly small since the approach starts by the candidate solution obtained by the Myopic approach and apply one pass of improvement instead of repeated ones in earlier steps.

Generally speaking, the new Hybrid approach provides better results than Neighborhood and Myopic approach, for all TSP-Lib instances listed in Table III. This means that the new Hybrid approach outperforms both the Neighborhood and Myopic approaches of the vertex *p*-median location problem.

2) Comparison vs. existing techniques: Table IV compared the performance of the Hybrid Myopic-Neighborhood versus the performance of two versions of Variable Neighborhood Search approach called Var1 and Var2 presented in [25]. The results are given in Table IV which shows the value of the objective function (*Z*), the deviations in % and the CPU time in seconds for the Hybrid approach. The notations in the table are the same as the ones given earlier for Table III.

To our surprise, the results demonstrated that the new Hybrid provides better solutions compared to Variable Neighborhood Search on all TSP-Lib instances listed in Table IV. The Hybrid approach yields new benchmarking solutions for all of the instances by producing 11 new best solutions which can be used for further benchmarking.

B. BIRCH Instances

BIRCH is a generated-synthetic dataset suggested by Zhang et al. [23]. Each BIRCH dataset contains *p* two-dimensional clusters demand points (data points) generated in a square. Dataset of type I is the easiest to solve while datasets II and III are harder [22]. Type 1 and Type 3 instances results are reported in this paper; these are the most frequently used types in the vertex *p*-median literature. The largest problem instance generated in this category contains 20,000 demand points. The number of medians (clusters) *p* is ranged from 25 and 100 as shown in Table V which also shows the results of the 24 BIRCH instances.

TABLE III. COMPUTATIONAL RESULTS FOR THE HYBRID, NEIGHBORHOOD “NBHD” AND MYOPIC APPROACHES ON THE TSP-LIB

n	***p***	***Z***			***Deviation (%)***		***Time (Sec)***		
		Hybrid	***NBHD***	***Myopic***	***NBHD***	***Myopic***	***Hybrid***	***NBHD***	***Myopic***
734	25	**205538.00**	213171.90	221036.85	3.71	7.54	1.4	0.02	0.47
929	30	**199171.23**	200646.89	208460.24	0.74	4.66	5.5	0.05	0.72
1979	35	**305849.50**	313868.19	329945.58	2.71	7.88	163.6	0.12	3.46
4663	40	**6508942.87**	6651807.60	6858853.82	2.14	5.38	376.0	1.04	44.62
6117	45	**2293933.42**	2388862.31	2483309.72	4.13	8.26	639.4	1.68	124.25
7146	50	**926303.48**	938372.66	982511.38	1.03	6.07	673.0	2.97	235.86
7663	55	**1337941.96**	1359608.60	1431263.27	1.41	6.97	1021.6	3.60	333.77
8246	60	**1206962.23**	1233756.02	1263392.14	2.29	4.68	1532.0	3.19	474.34
9847	65	**3211308.65**	3271312.72	3386102.36	1.99	5.44	2328.5	6.67	1134.28
9882	70	**1768824.70**	1780385.42	1892195.85	0.77	6.97	1958.3	8.52	1289.16
9976	75	**6172025.46**	6275422.16	6549036.63	1.72	6.11	2250.1	4.09	1161.07
Average					**2.06**	**6.36**			

TABLE IV. COMPUTATIONAL RESULTS FOR THE HYBRID, VAR1 AND VAR2 APPROACHES ON THE TSP- LIB INSTANCES

n	***p***	***Z***				***Deviation (%)***	
		Hybrid	***Var1***	***Var2***		***Var1***	***Var2***
734	25	**205538.00**	209214	207647		1.79	1.03
929	30	**199171.23**	208002	209374		4.43	5.12
1979	35	**305849.50**	320885	320777		4.92	4.88
4663	40	**6508942.87**	6885883	6817921		5.79	4.75
6117	45	**2293933.42**	2412269	2371727		5.16	3.39
7146	50	**926303.48**	1010761	1003380		9.12	8.32
7663	55	**1337941.96**	1430669	1416127		6.93	5.84
8246	60	**1206962.23**	1235810	1241036		2.39	2.82
9847	65	**3211308.65**	3365986	3311024		4.82	3.11
9882	70	**1768824.70**	1869116	1859538		5.67	5.13
9976	75	**6172025.46**	6378764	6340439		3.35	2.73
Average						**4.94**	**4.28**

1) Comparison vs. Neighborhood “NBHD” and Myopic: The computational results for our new Hybrid approach on the BIRCH dataset are presented in Table V, where the summary results of the three approaches (Hybrid, Neighborhood “NBHD” and Myopic) are shown. Here, we have the obtained objective function (*Z*) of the three approaches, the deviation (%) of the NBHD and Myopic form the Hybrid, and the run time (in seconds). On the BIRCH instances of Type 1 and Type 3, the Hybrid approach, as shown in earlier experiments, provides again better solutions compared to Neighborhood “NBHD” and Myopic approaches.

2) Comparison vs. existing techniques: The results of our computational experiments on the BIRCH I and III datasets are also compared with the results obtained by Hansen et al. [22], who presented a primal–dual variable neighborhood search (VNS) algorithm; and Avella et al. [24] who introduced a Lagrangean relaxation approach, which consists of three components: (1) subgradient column generation; (2) core heuristic; and (3) an aggregation procedure. The two approaches are referred to as VNS and CH respectively. The results of the VNS and CH approaches are taken from [24].

The computational results of the new Hybrid approach on the BIRCH dataset are presented in Table VI versus the results of the other two approaches (VNS and CH). For the BIRCH I instances the new Hybrid provides better solutions compared to VNS and CH. The Hybrid approach yielded the same results as the best-known results with deviation equal to (0.000), while VNS and CH yielded (0.039) and (3.674) respectively. On the BIRCH III instances, Hybrid outperforms VNS and CH where Hybrid found three best solutions and yielded the smallest deviation (0.002). As stated in [22] our experiments also show that the BIRCH instances of type 3 are harder to solve compared to type 1 instances. Also, Table VI shows that not only the Hybrid approach yielded better results than VNS and CH, but also requires relatively less computational burden compared to its nearest competitor in terms of quality, namely, the VNS metaheuristic. This reduction in computational time is even more significant in those larger instances where the hybrid consumes only approximately 15% of the time spent by the VNS.

TABLE V. COMPUTATIONAL RESULTS FOR THE HYBRID, NBHD AND MYOPIC APPROACHES ON THE BIRCH INSTANCES

Type 1		**Z**			***Deviation (%)***		***Time (Sec)***		
n	***p***	***Hybrid***	***NBHD***	***Myopic***	***NBHD***	***Myopic***	***Hybrid***	***NBHD***	***Myopic***
10000	100	**12,428.50**	12,428.50	15,587.56	0.00	25.42	3496.5	3.4	1920.1
15000	100	**18,639.30**	19,079.36	22,934.81	2.36	20.21	6515.9	12.9	4137.4
20000	100	**24,840.30**	25,462.10	30,931.91	2.50	21.48	7483.5	13.0	8648.0
9600	64	**11,934.80**	12,407.19	15,287.29	3.96	23.21	751.7	1.9	686.3
12800	64	**15,863.80**	15,863.81	20,005.21	0.00	26.11	1242.3	2.5	1382.2
16000	64	**20,004.60**	21,423.51	25,370.31	7.09	18.42	2135.9	6.8	1821.3
19200	64	**24,018.30**	24,964.97	31,225.47	3.94	25.08	2914.6	3.7	2711.1
10000	25	**12,455.70**	12,455.71	15,548.69	0.00	24.83	116.3	0.4	105.2
12500	25	**15,597.10**	15,597.15	19,734.60	0.00	26.53	1331.4	0.5	308.8
15000	25	**18,949.30**	18,949.25	23,531.83	0.00	24.18	413.1	0.9	261.6
17500	25	**21,937.40**	21,937.40	27,119.47	0.00	23.62	543.7	1.0	409.8
20000	25	**25,096.80**	25,096.82	31,082.81	0.00	23.85	897.9	0.8	558.7
Average					**1.65**	**23.58**			
Type 3		**Z**			***Deviation (%)***		***Time (Sec)***		
n	***p***	***Hybrid***	***NBHD***	***Myopic***	***NBHD***	***Myopic***	***Hybrid***	***NBHD***	***Myopic***
10000	100	**9,624.79**	10,023.04	10,542.17	4.14	5.18	2881.3	7.3	2737.2
15000	100	**15,904.12**	16,461.50	17,297.50	3.50	5.08	6252.1	11.4	5685.3
20000	100	**19,989.02**	20,757.61	21,958.13	3.85	5.78	7372.9	26.0	6783.1
9600	64	**8,225.58**	8,470.84	8,793.05	2.98	3.80	776.4	4.3	766.2
12800	64	**10,210.36**	10,597.02	11,779.22	3.79	11.16	1218.3	6.4	1193.5
16000	64	**13,340.47**	13,805.74	14,653.65	3.49	6.14	2337.1	12.1	1904.1
19200	64	**15,207.56**	15,671.28	16,915.79	3.05	7.94	3367.3	8.0	2606.2
10000	25	**7,203.39**	7,507.42	7,813.95	4.22	4.08	115.2	1.1	104.9
12500	25	**8,576.10**	9,219.43	10,033.18	7.50	8.83	278.5	1.5	171.9
15000	25	**9,513.64**	9,864.70	10,188.10	3.69	3.28	287.1	0.9	252.4
17500	25	**12,535.68**	13,686.14	14,877.32	9.18	8.70	465.5	1.4	347.2
20000	25	**13,052.81**	13,935.27	15,085.42	6.76	8.25	582.2	1.7	491.4
Average					**4.68**	**6.52**			

TABLE VI. COMPUTATIONAL RESULTS FOR THE VNS, CH AND HYBRID APPROACHES ON THE BIRCH INSTANCES

BIRCH instances of Type 1				***Deviation (%)***			***Time (Sec)***		
n	***p***	**Best-Known**		**VNS**	**CH**	**Hybrid**	**VNS**	**CH**	**Hybrid**
10000	100	**12428.5**		0.021	0.001	**0.000**	786	47	3,496.5
15000	100	**18639.3**		0.213	0.002	**0.000**	3,386	101	6,516
20000	100	**24840.3**		0.000	0.001	**0.000**	3,982	210	7,484
9600	64	**11934.8**		0.023	0.002	**0.000**	1,205	56	**752**
12800	64	**15863.8**		0.015	0.001	**0.000**	2,451	84	**1,242**
16000	64	**20004.6**		0.000	0.001	**0.000**	2,739	129	**2,136**
19200	64	**24018.3**		0.021	0.002	**0.000**	3,698	219	**2,915**
10000	25	**12455.7**		0.065	0.001	**0.000**	1,091	82	**116**
12500	25	**15597.1**		0.049	8.794	**0.000**	2,073	115	**1,331**
15000	25	**18949.3**		0.028	16.681	**0.000**	2,353	175	**413**
17500	25	**21937.4**		0.026	8.437	**0.000**	2,615	241	**544**
20000	25	**25096.8**		0.001	10.168	**0.000**	3,055	365	**898**
Average				0.039	3.674	**0.000**	2,453	152	**2,320**
BIRCH instances of Type 3				***Deviation (%)***			***Time (Sec)***		
n	***p***	**Best-Known**		**VNS**	**CH**	**Hybrid**	**VNS**	**CH**	**Hybrid**
10000	100	**9624.79**		0.096	0.002	**0.002**	2609	60	2,737.2
15000	100	**15904.12**		0.094	21.767	**0.005**	3,495	121	5,685
20000	100	**19989.02**		0.181	27.983	**0.003**	3,429	222	6,783
9600	64	**8225.58**		0.123	21.912	**0.002**	1,483	57	**766**
12800	64	**10210.36**		0.117	11.412	**0.001**	2,503	98	**1,194**
16000	64	**13340.47**		1.890	23.142	**0.001**	3,169	170	**1,904**
19200	64	**15207.56**		0.907	38.925	**0.006**	3,243	229	**2,606**
10000	25	**7203.39**		0.834	11.349	**0.000**	1,016	94	**105**
12500	25	**8576.1**		0.788	0.956	**0.000**	1,606	144	**172**
15000	25	**9513.64**		3.099	52.041	**0.003**	2,742	192	**252**
17500	25	**12535.68**		1.141	38.387	**0.000**	2,803	250	**347**
20000	25	**13052.81**		2.060	54.700	**0.003**	3,364	364	**491**
Average				0.944	25.215	**0.002**	2,622	167	**1,920**

VI. Conclusion

This paper introduces a new Hybrid heuristic to solve large scale vertex *p*-median location problems varying in size ranging from 734 to 20,000 demand points. Many Heuristics, Meta-heuristics and Exact approaches have been developed for this purpose. This paper presented a new Hybrid-heuristic approach which integrates two heuristic approaches, namely, the Myopic approach as a construction method to find the solution of each q ($q=1,..,p$) median facility; while the Neighborhood approach improves this solution as much as possible at each level of q. By embedding the Neighborhood approach into the Myopic heuristic within the search and not put in a sequential manner as a post optimizer at the very end, as usually applied, excellent results have been produced which are highly competitive with the state-of-the-art approaches on large-size instances with up to 20,000 demand points. This can be seen to act as a continuous filtering mechanism to guide the search.

The new Hybrid approach was tested on the TSP-Lib instances (n = 734−9976) and outperformed the ones by [25]. In addition, the new approach was assessed on several large-size BIRCH instances (n = 9600−20000), each instance is solved with p ranging from 25 to 100. The results show that our method gives better solutions compared to [24] and [22] results.

In brief, the results show that the new approach gives in general better solutions which can be then used for benchmarking purpose in the future. This demonstrates that a simpler approach that takes into account the advantages of other methods can lead to promising outcome besides having the potential to be adopted in tackling other combinatorial optimization problems.

Acknowledgment

I would like to acknowledge Prophet Muhammad (PBUH) who has taken all human being from the darkness to the lightness.

References

[1] Irawan, S. Salhi, and Z. Drezner, "Hybrid meta-heuristics with VNS and exact methods: application to large unconditional and conditional vertex p-centre problems " Journal of Heuristics, 2015.

[2] H. M. Rabie, I. A. El-Khodary, and A. A. Tharwat, "Particle Swarm Optimization algorithm for the continuous p-median location problems," in 2014 10th International Computer Engineering Conference (ICENCO), 2014, pp. 81-86: IEEE.

[3] J. Janáček and M. Kvet, "Sequential approximate approach to the p-median problem," Computers & Industrial Engineering, vol. 94, pp. 83-92, 2016.

[4] C. Öztürk, G. Tuzkaya, and S. Bulkan, "Centrality based solution approaches for median-type incomplete hub location problems," Computers & Industrial Engineering, vol. 156, p. 107275, 2021.

[5] H. M. Rabie, "Particle swarm optimization and grey wolf optimizer to solve continuous p-median location problems," in Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges: Springer, 2021, pp. 415-435.

[6] S. L. Hakimi, "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph," Operations Research, vol. 12, no. 3, pp. 450-459, 1964.

[7] S. L. Hakimi, "Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems," Operations Research, vol. 13, no. 3, pp. 462-475, 1965.

[8] O. Kariv and S. L. Hakimi, "An algorithmic approach to network location problems. Part 2:The p-Medians," SIAM Journal of Applied Mathematics, vol. 37, no. 3, pp. 539–560, 1979.

[9] C. A. Irawan, S. Salhi, and M. P. Scaparra, "An adaptive multiphase approach for large unconditional and conditional p-median problems," European Journal of Operational Research, vol. 237, pp. 590–605, 2014.

[10] H. M. Rabie, "Particle Swarm Optimization and Grey Wolf Optimizer to Solve Continuous p-Median Location Problems," Cham, 2020, pp. 136-146: Springer International Publishing.

[11] R. Z. Farahani and M. Hekmatfar, Facility Location Concepts, Models, Algorithms and Case Studies. Springer-Verlag Berlin Heidelberg, 2009.

[12] Z. Drezner, J. Brimberg, N. Mladenović, and S. Salhi, "New Heuristic Algorithms for Solving the Planar p-Median Problem," Computers & Operations Research, 2014.

[13] R. Z. Farahani, N. Asgari, N. Heidari, M. Hosseininia, and M. Goh, "Covering problems in facility location: A review," Computers & Industrial Engineering, vol. 62, pp. 368–407, 2012.

[14] C. S. ReVelle and R. W. Swain, "Central Facilities Location," Geographical Analysis, vol. 2, no. 1, pp. 30-42, 1970.

[15] K. Rosing, C. ReVelle, and H. Rosing-Vogelaar, "The p-median model and its linear programming relaxation: an approach to large problems," The Journal of the Operational Research Society, vol. 30, no. 9, pp. 815–823, 1979.

[16] P. Avella, A. Sassano, and I. Vasil'ev, "Computational study of large-scale p-median problems," Math. Program, vol. 109, pp. 89–114, 2007.

[17] H. M. Rabie, I. A. El-Khodary, and A. A. Tharwat, "A particle swarm optimization algorithm for the continuous absolute p-center location problem with Euclidean distance," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 4, no. 12, pp. 101-106, 2013.

[18] S. Salhi, The Palgrave Handbook of Operations Research. Springer Nature, 2022.

[19] G. Laporte, S. Nickel, and F. Saldanha-da-Gama, "Introduction to location science," in Location science: Springer, 2019, pp. 1-21.

[20] G. Laporte, S. Nickel, and F. S. d. Gama, Location Science. Springer 2015.

[21] G. Reinelt, "TSLIB—a traveling salesman library," ORSA J. Comput, vol. 3, pp. 376–384, 1991.

[22] P. Hansen, J. Brimberg, D. Urošević, and N. Mladenović, "Solving large p-median clustering problems by primal–dual variable neighborhood search," Data Mining Knowledge Discovery, vol. 19, no. 3, pp. 351-375, 2009.

[23] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: A new data clustering algorithm and its applications," Data Mining Knowledge Discovery, vol. 1, no. 2, pp. 141-182, 1997.

[24] P. Avella, M. Boccia, S. Salerno, and I. Vasilyev, "An aggregation heuristic for large scale p-median problem," Computers & Operations Research, vol. 39, pp. 1625–1632, 2012.

[25] C. A. Irawan, S. Salhi, and M. P. Scaparra, "An adaptive multiphase approach for large unconditional and conditional p-median problems," European Journal of Operational Research, vol. 237, no. 2, pp. 590-605, 2014.

[26] C. A. Irawan and S. Salhi, "Solving large p-median problems by a multistage hybrid approach using demand points aggregation and variable neighbourhood search," Journal of Global Optimization, vol. 63, no. 3, pp. 537-554, 2015.

[27] K. E. Rosing and C. ReVelle, "Heuristic concentration: Two stage solution construction," European Journal of Operational Research, vol. 97, no. 1, pp. 75-86, 1997.

[28] I. Vasilyev and A. Ushakov, "A shared memory parallel heuristic algorithm for the large-scale p-median problem," in International Conference on Optimization and Decision Science, 2017, pp. 295-302: Springer.

[29] I. Vasilyev, A. V. Ushakov, N. Maltugueva, and A. Sforza, "An effective heuristic for large-scale fault-tolerant k-median problem," Soft Computing, vol. 23, no. 9, pp. 2959-2967, 2019.

[30] M. S. Daskin, Network and Discrete Location: Models, Algorithms, 2nd Edition ed. John Wiley and Sons, Inc., New York., 2015.

[31] L. Cooper, "Heuristic methods for location-allocation problems," SIAM review, vol. 6, no. 1, pp. 37-53, 1964.

[32] S. Salhi and R. Atkinson, "Subdrop: A modified drop heuristic for location problems," Location Science, vol. 3, no. 4, pp. 267-273, 1995.

Appendix A

For the first median (p=1) Set X=Ø, Table 7 shows the calculations of the total enumeration by summing the entries in each column in Table 1, to obtain the values of $Z(J,X)$. Finding i^* =argmin$_{i\epsilon I}$ { $Z(J,X)$}. The smallest value is 4772, i^*=9 and X={9}.

TABLE VII. The Total Enumeration (P=1)

I	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Z	**7816**	**7913**	**5855**	**6457**	**5784**	**5760**	**6936**	**5971**	**4772**	**6886**	**5576**	**6371**

For the second median (p=2), (1) applying Myopic approach by finding i^* =argmin$_{i\epsilon I}$ {Z(J, X U{i})}. Table 8 shows the results of this computation, to obtain the values of $Z(J,X)$. The minimum sum value is 3145, corresponds i^*=7 and X={9,7}.

TABLE VIII. The Results of the Myopic Approach (P=2)

X	9	9	9	9	9	9	**9**	9	9	9	9	9
i	1	2	3	4	5	6	**7**	8	9	10	11	12
Z	3845	3725	3943	4296	3696	4268	**3145**	3655	4772	3563	3858	4392

According to the Myopic, we should proceed by adding a new third point (vertex/median) in the same manner, however, the new Hybrid Myopic-Neighborhood approach proposed applying the Neighborhood approach. By finding N_1 and N_2, where N_i is the set of demand nodes which i median is the closest open median facility to it. Here $N1$ ={3,4,5,6,9,11,12} and $N2$ ={1,2,7,8,10}. Then find k^* =argmin$_{k\epsilon Ni}${$Z(Ni$,{k})}; k^*=9,7 and X^{new}={9,7}. No median facility changed; Go to Next median (p).

For the third median (p=3), (1) applying Myopic, by finding i^* =argminiϵI {$Z(J$, X U{i})}. Table 9 shows the results of this computation to obtain the values of $Z(J,X)$. The minimum value is 2641, corresponds i^*=6 and X={9,7,6}.

TABLE IX. The Results of the Myopic Approach (P=3)

X	9	9	9	9	9	**9**	9	9	9	9	9	9
	7	7	7	7	7	**7**	7	7	7	7	7	7
i	1	2	3	4	5	**6**	7	8	9	10	11	12
Z	2695	2770	2647	2689	2929	**2641**	3145	2845	3145	2661	2718	2765

(2) applying Neighborhood, by finding N_1, N_2 and N_3. Here N_1 ={3,5,9,11,12} N_2 ={1,2,7,8,10}, and N_3 ={4,6}. Then find k* =argmin$_{k\epsilon Ni}${$Z(N_i$,{k})}; k^*=11,7,6 and X^{new}={11,7,6}, the new value of Z is 2535. Median facility changed; go to step 4.

finding N_1 , N_2 and N_3. N_1 ={8,10,11,12}, N_2 ={1,2,5,7} and N_3 ={3,4,6,9}. Then find k^* =argmin$_{k\epsilon Ni}${$Z(N_i$,{k})}; k^*=11,1,6 and X^{new}={11,1,6} the new value of Z is 2438. Median facility changed; go to step 4.

finding N_1 ,N_2 and N_3. N_1 ={8,10,11,12} N_2 ={1,2,5,7} and N_3 ={3,4,6,9}. Then find k^* =argmin$_{k\epsilon Ni}${$Z(N_i$,{k})}; k^*=11,1,6 and X^{new}={11,1,6} No median facility changed; Go to Next p median and so on.

Annotated Corpus with Negation and Speculation in Arabic Review Domain: NSAR

Ahmed Mahany[1]*, Heba Khaled[2], Nouh Sabri Elmitwally[3], Naif Aljohani[4], Said Ghoniemy[5]
Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt [1, 2, 5]
School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, UK[3]
Faculty of Computers and Artificial Intelligence, Cairo University, Giza 12613, Egypt[3]
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia[4]

***Abstract*—Negation and speculation detection are critical for Natural Language Processing (NLP) tasks, such as sentiment analysis, information retrieval, and machine translation. This paper presents the first Arabic corpus in the review domain annotated with negation and speculation. The Negation and Speculation Arabic Review (NSAR) corpus consists of 3K randomly selected review sentences from three well-known and benchmarked Arabic corpora. It contains reviews from different categories, including books, hotels, restaurants, and other products written in various Arabic dialects. The negation and speculation keywords have been annotated along with their linguistic scope based on the annotation guidelines reviewed by an expert linguist. The inter-annotator agreement between two independent annotators, Arabic native speakers, is measured using the Cohen's Kappa coefficients with values of 95 and 80 for negation and speculation, respectively. Furthermore, 29% of this corpus includes at least one negation instance, while only 4% of this corpus contains speculative content. Therefore, the Arabic reviews focus more on negation structures rather than speculation. This corpus will be available for the Arabic research community to handle these critical phenomena[1].**

Keywords*—Arabic NLP; negation; speculation; uncertainty; annotation; annotation guidelines; corpus; review domain; sentiment analysis

I. Introduction

Negation and speculation are commonly used linguistic phenomena, providing information on factuality and the polarity of facts [1]. Negation is a linguistic property shared by all human languages [2], which denotes the absence of something; therefore, negation affects the contextual polarity of words. On the contrary, speculative language is used to convey uncertainty about an event or idea. It means there is not enough evidence in the text to prove whether the information is 100% true. Consequently, sentences including negation or speculation may misclassify the opinionated phrases [3] or inaccurately identifying the medical terms [4], [5]. In order to efficiently identify instances of these phenomena, it is necessary to find those words expressing negation and speculation and then their scope, such as the tokens within the sentence that are affected by these cues [6]. Since negation and speculation are language-dependent, they must be addressed in all-natural languages [7]. Therefore, many studies addressed them to enhance the performance of Natural Language Processing (NLP) tasks and applications in various languages such as Sentiment Analysis (SA) [8], Machine Translation (MT) [9], and Information Extraction (IE) [5]. These studies addressed the negation and speculation scope detection using rule-based [10] and sophisticated supervised learning methods [11], [12].

Arabic Natural Language Processing (ANLP) has gained unprecedented interest in the age of big data and social media platforms, making it one of the most important research topics, especially in North Africa and the Gulf Area [13]. Classical Arabic (CA), Modern Standard Arabic (MSA), and Dialectal Arabic (DA) are the three primary forms of Arabic [14]. The Qur'an and ancient literature are written in the CA form. The MSA is mainly used in education, the official written reports like newspapers, and formal TV programs. Conversely, the DA includes all current forms of Arabic spoken, written on social media platforms, and reviewed applications and websites where it varies nationally and internationally depending on location [15]. Since the DA has no syntactic rules and multiple forms of the same word, ANLP tasks are challenging.

Negation frequently occurs in the Arabic language and is one of the dominant linguistic methods for changing the text polarity, so negation detection is highly considered in the Arabic Sentiment Analysis (ASA) [3]. However, the presence of negation words in a sentence does not imply that all the sentimental words are inverted. Still, there are odd cases where the presence of negation terms may confirm the polarity of the following lexeme [16]. In the implicit form of negation, a sentence can be negated without using negation words. The level of speculative content increases or decreases the certainty of polarity classification [17]. Few Arabic studies have addressed the impact of negation and speculation using simple rules. Hamouda and El-Taher considered the frequency of negation terms in the ASA task as a classification feature, but the effect of the negation feature on the sentiment classification was not clearly mentioned [18]. In 2015, Duwairi and Alshboul defined six handcrafted rules to handle negation in the Modern Standard Arabic (MSA) texts in the review domain to enhance the performance of the ASA [19]. Even though they addressed the MSA, which follows well-defined rules, the simplistic approach has proven inadequate for a syntactically and morphologically rich language like Arabic. El-Naggar et al. considered several valences to build a negation-aware classifier for ASA in MSA and the Egyptian dialect [20]. Later, Assiri et al. formulated four rules to handle negation in the Saudi dialect

*Corresponding Author

[1] https://github.com/amahany/NSAR

[21]. In addition, Kaddoura et al. have proposed a system that inverts the polarity of a sentence's clause if a negation term precedes a positive or negative pattern [3]. Regardless of the improvement in performance in these systems' experimental results [3], [20], [21], none handled the implicit form of negation frequently used in Arabic. Simple rule-based algorithms cannot handle all the negation and speculation cases for the various Arabic language forms and dialects [14]. According to the findings of our earlier work, the treatment of negation scope detection utilizing supervised based learning is promising [12]. To the best of our knowledge, there are no available Arabic corpora annotated with negation or speculation in various domains including the review, newswire and medical domains. Furthermore, speculation detection in ASA has not been studied in any research work.

In the last decade, there has been a growing interest in detecting negation and speculation. Nevertheless, the available open-access corpora for low-resource languages, such as the Arabic language [22], are limited compared to the English and the Spanish languages [7]. Speculation corpora are even more scarce than those for negation, with the majority focusing on the biomedical domain. Since negation and speculation are language-dependent phenomena, the negation- and speculation-aware models from other languages, such as English, cannot be applied to the Arabic text because the syntactic structure of negation in Arabic differs from that in English. Therefore, developing an annotated corpus with negation and speculation for the Arabic review domain is required. It is very important to know that negation- and speculation-aware systems improve the overall systems performance [9], [11].

The rest of the paper is organized as follows: Section II shows the different sources for our corpus. Section III details the annotation guidelines we build for the negation and speculation texts in the Arabic review domain. The annotation process and its result including the agreement analysis of the annotators and the discussion are presented in IV and V. Finally, Section VI concludes the paper and suggests the future work.

II. Corpus Collection

This section demonstrates the overall characteristics of the Negation and Speculation Arabic Review (NSAR) corpus, as well as a brief description of the texts that compromise it. Furthermore, general statistics are presented regarding each source's size and polarity distribution. The NSAR corpus is comprised of texts extracted from three well-established and benchmarked Arabic review corpora: Large Scale Arabic Book Review (LABR) [23], Large Arabic Multi-domain Resources (LAMR) [24], and Multi-domain Arabic Sentiment Corpus (MASC) [25]. Table I shows the distribution of randomly selected positive and negative sentences from each source, with 2,312 positive reviews accounting for approximately 77% of our corpus. Each topic has a different number of sentences, but the average number of words per sentence is nearly the same. The LABR corpus contains 63K book reviews, with ratings ranging from 1 to 5 stars [23]. Aly and Atiya considered the reviews with 4 or 5 stars with positive polarity and those with 1 or 2 stars with negative polarity. The authors collected these reviews from the best Arabic books listed in the social network for book readers [2]; hence, most of the randomly selected reviews are positive reviews, as per Table I. The LAMR corpus is the second source for NSAR corpus, and it consists of 33K reviews scrapped via Scrapy framework [3] from various reviewing websites, Souq [4], TripAdvisor [5], Elcinema [6], and Qaym [7], including reviews for various items and services [24]. Each sentence includes the review text and normalized rating that could be positive, negative, or mixed polarity. The third source, MASC [25], includes 8,860 reviews on different topics such as shopping, restaurants, and software applications written in multiple Arabic dialects. These reviews were obtained primarily from Jeera [8], Qaym, Google Play, Twitter, and Facebook. The majority of the reviews from LAMR and MASC were composed in Egyptian, and Gulf areas' dialects. On the contrary, most of the LABR samples were written in the MSA form. The review texts in the NSAR corpus are collected from various sources to ensure that it captures the diversity of dialectical language usage in the review domain.

TABLE I. Corpus Statistics

Corpus	Topic	Positive	Negative	Total
LABR	Books	879 (84.52%)	161 (15.48%)	1,040
LAMR	Touristic Attractions	102 (94.44%)	6 (5.56%)	108
	Hotels	74 (100%)	0 (0%)	74
	Products	684 (75.58%)	221 (24.42%)	905
	Resturants 1	248 (74.47%)	85 (25.53%)	333
	Restaurants 2	114 (81.43%)	26 (18.57%)	140
MASC	Software	210 (52.50%)	190 (47.50%)	400
	Products	1 (9.09%)	10 (90.91%)	11
Total		**2,312 (76.79%)**	**699 (23.21%)**	**3,011**

III. Annotation Guidelines

Negation and speculation phenomena are interrelated and have similar characteristics: they both have a scope, so they affect the part of the text denoted by the presence of negating or speculative keywords. Furthermore, both of them have two types: implicit and explicit. In the case of the explicit type, the phenomenon cue is written in the sentence, whereas being understood in the case of the implicit one without a cue. Sentences including a negation cue are not necessarily annotated for negation; however, they may have speculative content. Therefore, the annotators should read sentences containing negation cues carefully. In most cases, the keywords influence their scope, aligned from the left to the end of the clause or the sentence.

The following subsections list the general principles, negation, and annotation guidelines. Furthermore, the special

[2] https://www.goodreads.com/
[3] https://scrapy.org/
[4] https://souq.com/
[5] https://www.tripadvisor.com/
[6] https://elcinema.com/
[7] https://www.qaym.com/
[8] https://www.jeeran.com/

or complex cases for both phenomena are demonstrated. In order to illustrate examples in the annotation guidelines, the negating cues are surrounded by a negation symbol (¬), the speculative cues are surrounded by an uncertainty symbol (∓), and their scope boundaries are surrounded by parenthesis.

A. General

When annotating the negation and speculation, several general rules must be followed, which are adapted from the BioScope annotation guidelines [6], then modified to the Arabic language and review domain. Sentences with some instance of negative or speculative language will be only considered. In addition, the min-max strategy should be followed during the annotation. The minimal unit (single word) that expresses the negation or speculation will be marked as a cue. Nevertheless, in some cases, a cue may include more than a single word which is called a complex cue. The maximum number of words affected by a cue will be marked as the scope for negation or speculation. The scope usually starts after the keyword and ends at the end of the phrase, clause, or sentence. However, the scope may include a word or a statement preceding it. The below list summarizes the general rules for both negation and speculation:

- A sentence may contain more than one cue instead of only one keyword; in this situation, each cue should be annotated separately.
- Structures of negation and speculation can be annotated in a single sentence.
- The cue is not included in the scope, but it may be included in complex cases and in the scope that includes words preceding and following a cue.
- If a sentence contains a cue that appears at the end of the sentence, the phenomenon's scope is limited to the cue.
- Due to the improper use of spaces in the informal Arabic text, a cue+verb/noun may be concatenated without a space; in this case, the verb/noun will be included in the negation/speculation cue.
- The coordinating conjunctions و (and) extend the scope.
- Annotators will only annotate the cue and leave the scope for the linguist expert if the annotator is unsure about the scope.
- There is an annotation element called the ‘undecided’ used if the annotator is unsure what type the keyword should be assigned.

Additionally, each type of a negation or speculation structure is depicted with an example where the transliteration and English translation of these examples are listed in Appendix I.

B. Negation Structures

- لا (*no*) is the most used negating Arabic word, which is used to deny the occurrence of a verb in the past and present tenses, as well as to deny a nominal sentence. Therefore, the scope begins with the negative cue and ends at the end of the sentence.

1) أحلام اجمل الايام كانت هناك و (ذكريات ¬ لا ¬ تنسى) مكان رائع يجمع الكل فى حركه و ضحك وحياه

- ما (*did not – does not*) interfere with past or present verbs in dialectal Arabic and the formal Arabic forms in the nominal sentences. It is most often found in a pre-verbal state. To ensure that ما with a verb in the past is working, replace it with لم (*does not*) followed by the same verb in the present.

2) لو كان سعره اقل وفيه فلاش كنت قيمته اكار من كدا فشل واضح من سوني انها ¬ ماحطتش ¬ (ليه فلاش)

- لم keyword is used with a verb in the present tense to deny a truth that occurred in the past. In exceptional cases, it may affect something in the present or future.

3) جهاز جبار ويتفوق علي نظرائة من الايباد والسامسونج بجد رائع بس ¬ لم ¬ (ياخذ حجم الدعايه المطلوبه)

- لما (*Lamma*) is used before a verb in the present to deny something in the past and may deny something in the future. However, if it comes with a verb in the past, it does not deny anything after it.
- لات (*no*) cue is used only in the classical Arabic form.
- إن (*Inn*) affects nouns, past, or present verbs. إن will be effective if it gets replaced with another cue and reverses the polarity. There is a distinction between إن و أن (*Ann*) where أن is not a negating cue. Furthermore, إن و إن may be written ان without همزة (*Hamza*) in the dialectical Arabic. Therefore, it is necessary to read the sentences carefully to determine the correct form in accordance with the context of the sentence.
- لن (*will not*) is used with a verb in the present tense to deny something in the future.

4) 7 روايات في رواية ، تهكم وسخرية وأدب وأشياء اخرى ¬ لن ¬ (تمل منها)

- ليس (*Not*) is a verb in the past tense, extending the negation to the end of the sentence. The origin of ليس is لا+ايس which means no + existing. It has different forms like لست – ليست – ليسا – لسنا – لسن – فلست

5) يستاهل لانه دعم اللغة العربية لكن ياريت يثبت البرنامج على رقم المستخدم ¬ وليس ¬ (على حساب)

- عدم (*None*), عدا (*Except*), and دون (*Without*) keywords are used to deny a noun or nominal sentence.

6) الذي فاجئني وزعجني هو ¬ عدم ¬ (وجود ريموت فيها) بالمقارنة مع السعر

- مش (*Mesh*) keyword and the pattern of م + verb + ش are mainly used in the Egyptian dialect to deny a verb.

7) تطبيق رائع وفيه خصوصيه ¬ محدش ¬ (يعرف دخلت امتي خرجت امتي) انت حتي لو بتكتب (اللي قدامك ¬ مش ¬ بيعرف)

- مفيش (*Not*) used in the Egyptian dialect to deny nouns

8) فكرني بفلاش و سماش والحاجات دي بس ¬ مفيش ¬ (كلمات متقاطعة (

- مو (*Not*) keyword used in the Gulf/Levant countries to deny a verb, noun, or an event.

9) مرررررره حلو وعصيراته فرش ⌐ ومو ⌐ (حاطين له لاسكر ولا مويه كله فرش)

C. Speculation Structures

- There are many Arabic adjectives or adverbs that indicate speculation such as, محتمل – احتمال – ممكن – مرجح – شكاك – أحيانا –

10) نصيحتي ان (في محلات فطاير تانية في حدايق حلوان ∓ ممكن ∓ تكون أفضل كتير)

- Some other adjectives, if get preceded by a negation cue, it indicates a speculative content, such as غير مؤكد

11) اول روايه اقرأها لباولو كويلهو ∓ ولااعتقد ∓ (انها الاخيره)

- This list of verbs, but not limited, indicates speculation - يعتقد - يظن - يمكن – توحى – يبدو – تظهر – تشير – يفترض – يشك. In addition to, the noun form for some of these verbs like الشك – الظن – فرضية – الافتراض

12) ⌐ لم ⌐ (استسيغ نزار فى شعر الفصحى) ∓ اعتقدوا ∓ فى الشعر الحر اكثر ابداعا)

- These Arabic particles لو – قد – ربما – لعل –ما بين indicate speculative content.

13) فندق مريح صراحة اسعارة جيدة ∓ ما بين ∓ (150 الى 300 ريال الليلة) ونظيف جدا

- Conjunction keywords such as أو (*or*) have the scope of elements ranging from the right to the left side of the conjunction. However, in instances where the conjunction is composed of two or more words like أو, إما (*Or*), سواء (*Whether*), the scope does not change.

14) أنصح بيه أي حد بيدرس (هندسة كهربية ∓ أو ∓ عايز يدرسها) هتوفر عليه جنون كتير

- Sentences starting with a question that should be annotated as speculative.

- If the speculation cue is present at the start of the sentence, then the scope extends to include the whole sentence.

15) ∓ربما ∓ (يكون الكتاب جيداً ولكن بروز شخصية الكاتب المتملقه تفسد ذلك؟)

D. Negation Complex Cases

The presence of a negation keyword does not automatically negate a sentence as follows:

- For example, إن that assures something.

16) ⌐ ما ⌐ إن رأيت ولا سمعت بمثله

- Example for ما used for wonderment.

17) ⌐ ما ⌐ هذه الرومانسية الحالمة ⌐ وما ⌐ هذا الاسلوب الناعم الجميل هذه الرواية من اجمل ⌐ ما ⌐ قرات على الاطلاق

- Question marked withليس reverts the sentence from being negated to being proven

18) ⌐ أليس ⌐ هذا بالحق

- If a sentence consists of ما then إلا, the negation is canceled

19) ⌐ مَا ⌐ هَٰذَا بَشَرًا إِنْ هَٰذَا إِلَّا مَلَكٌ كَرِيمٌ

- Two consecutive negating cues like ما cancels the effect of negation.

- غير sometimes used in the Gulf countries to express something unique.

20) جدة ⌐ غير ⌐

- غير in some cases means change but not a negation cue

21) كتاب ⌐ غير ⌐ مجرى تفكيرى خلاه اوسع خلانى أثق اوووى فى العلامات

- مش بس is used to assure something

22) تجنننننن انا بصراحه ⌐ مش ⌐ بس الفندق جميل كل حاجه زورتها كانت جميله قوي قوي قوي قوي

- Verb in forms of ما + أفعل

23) أجمل ⌐ ما ⌐ فيه هو إفطاره

In some other cases, the negation is implied in the sentence without any negating cue while understood from the context of the text.

- The sentence implies denial without any negative cues such as

24) كتاب خفيف و واقعي و ⌐ بعيد ⌐ (عن المبالغة تماما) كل شيء فيه حقيقي

- Negating the not obvious

25) (وَتَرَى النَّاسَ سُكَارَىٰ وَمَا هُم بِسُكَارَىٰ)

E. Speculation Complex Cases

Certain speculation cases are marked using few keywords.

- For example, قد can express speculative content only if the verb following it is in the present tense form, as in point 25. However, the content in point 26 is not speculative and comes with a verb in the past tense.

26) الكلمة ∓ قد∓ (تفعل فى الانسان ما لم تفعلة الادوية القوية) لك كل قدير

27) السلعة ليست بالجودة المطلوبة ∓ وقد ∓ اشتخدمتها لمرة واحدة فقط ولم ارجع لاستخدامها مرة ثانية

- When used at the end of a sentence, the negation cue ما indicates speculation.

28) جلسات رائعة جلسات المطعم الخارجية رائعة خصوصا في فصل الربيع والشتاء اما (الاكل فجيد) ∓ نوعا ما ∓

- Most of the cases that use لعل do not express speculation; however, it represents hopefulness.

29) وَجَعَلَ لَكُمُ السَّمْعَ وَالْأَبْصَارَ وَالْأَفْئِدَةَ ۙ لَعَلَّكُمْ تَشْكُرُونَ

- In some cases, speculation cues may be used to imply an affirmation.

30) الَّذِينَ يَظُنُّونَ أَنَّهُم مُّلَاقُو رَبِّهِمْ وَأَنَّهُمْ إِلَيْهِ رَاجِعُونَ

IV. NSAR Annotation

This section describes the procedure followed in the annotation process of the NSAR corpus. Initially, the guidelines are created based on the negation rules of the formal Arabic language in addition to the commonly used slang negating cues in the Egyptian and Gulf countries' dialects. Then, a list of Arabic keywords for the speculation is built which would indicate speculative content, and subsequently, these rules are applied to annotate a sample of the corpus and extract any additional cases from the corpus to enhance these rules for the annotation process.

There is a need for a tool for the annotation process to build and develop NSAR corpus. There are many available annotation tools for this purpose. Based on an evaluation of the well-known annotation tools in this study [26], WebAnno[9] is selected, which achieved the highest score [27]. WebAnno is an open-source web-based annotation tool that provides full functionality for both semantic and syntactic annotations. Furthermore, it supports adding user-defined annotation layers as we did for the negation and speculation. The user-defined layers are only supported in TSV3 format, where there is an open-source Python library to extract the annotations written in TSV[10]. As in Section II, NSAR corpus is collected from three different Arabic corpora from the review domain labeled as positive or negative and written in CSV file format. Therefore, we transformed the input files from CSV to TSV file format. Five user-defined labels associated with the WebAnno project: sentiment, negation, speculation, bad, and undecided are created. The sentiment has one feature called ‘polarity’ with ‘negative’ or ‘positive’ values, used with the transformation from CSV to TSV for the sentiment labeling. For the negation and speculation labels, every label has a tag set with two different values ‘cue’ and ‘scope’ which are associated to each other using two user-defined relations ‘NegRel’ and ‘SpecRel’. The other two labels ‘bad’ and ‘undecided’ are used to highlight any inappropriate or hateful content in the text or the annotator cannot take a decision about a sentence.

The annotation process was implemented in three phases: the first phase was to describe the annotation guidelines and train the annotators on using WebAnno, then the annotators carried out the annotation to measure the inter-annotator agreement (IAA), and finally, a linguist expert resolved the disagreements between them. Two independent Arabic native speakers carried out this process; one is an experienced annotator with a solid background, and the second is a well-trained person. Each file has been annotated by both annotators.

V. Results and Discussion

In this section, we explore the result of the annotation process. The Cohen’s Kappa coefficient [28] is used to measure the quality of the annotation process. Cohen’s Kappa of value 0.95 for the negation and 0.8 for speculation are obtained. These values demonstrate that the speculation annotation is more complex than the negation in Arabic. Table II shows the NSAR corpus, which includes 862 negated sentences out of 3,011, and only 121 sentences containing at least one speculative content.

The disagreements between the two annotators were revised by a linguist expert [6]. The majority of disagreement cases in negation are caused by common human errors, such as one of the annotators forgetting to relate the negation cue to its scope using the relation layer. Since a single sentence may contain multiple negation structures [29], this layer is added and should be specified for each annotation. The speculation cases, on the contrary, are ambiguous and may lead the annotator to consider it a negation or speculation [7]. Therefore, it had a higher level of disagreement than the negation. These cases involve an issue within the scope of speculation, such as the non-inclusion of a word. In addition to the undecided label, the disagreements have been curated by the first author and the linguist expert.

Table II shows that 29% and 4% of total sentences have at least negation and speculation structures, respectively; however, these percentages vary from topic to topic. For instance, MASC sub-corpus includes high rates of negating and speculative content.

TABLE II. NSAR Statistics

Corpus	Topic	Size	Negation	Speculation
LABR	Books	1,040	248 (23.85%)	46 (4.42%)
LAMR	Touristic Attractions	108	20 (18.52%)	1 (0.93%)
	Hotels	74	7 (9.46%)	2 (2.70%)
	Products	905	284 (31.38%)	30 (3.31%)
	Resturants 1	333	98 (29.43%)	10 (3%)
	Restaurants 2	140	33 (23.57%)	3 (2.14%)
MASC	Software	400	166 (41.50%)	28 (7.00%)
	Products	11	6 (54.55%)	1 (9.09%)
Total		**3,011**	**862 (28.63%)**	**121 (4.02%)**

The subject types in Arabic sentences change the form of most Arabic words, such as verbs ذهب (He went) and ذهبت (She went). There are other various forms of negation in Arabic that have the same meaning in English. This example shows the negation difference between the MSA and Egyptian dialect where ملكشى in the Egyptian dialect is derived from لا شىء لك or ليس لك شىء in MSA form, where all of them means (you do not own anything). Another example, مكنتش in the Egyptian dialect, which is derived from لم تكن or لم أكن in MSA, means (I do not + verb) or (She does not + verb) according to the context. However, removing a single character from this word as مكنش will change the meaning to be (He does not + verb). These examples demonstrate the complexity of negation in Arabic, especially in the dialect Arabic. Furthermore, the spelling rules are not followed in dialectical Arabic, resulting in tokenization issues such as in الكتابةلاتظهر (The written text does not appear) [3]. There is no space between the three words that should formally be used. Other instances in the dialect of Arabic include different forms for the same Arabic word with the same meaning as in مافيش and مفيش (None-existence). Therefore, we normalized the commonly used negation and

[9] https://webanno.github.io/webanno/

[10] https://github.com/neuged/webanno_tsv

speculation cues, as depicted in Table III and Table IV. The Negator لا and speculative cue لو account for approximately 45% of the negation and speculation cues, respectively.

TABLE III. THE COMMON NEGATING CUES IN NSAR

Normalized Negation Cues	Frequency
لا	455
ما	161
لم	129
غير	84
مش	78
لن	20
دون	25
مو	30
ليس	74
عدم	21
عدا	7
مفيش	5
ملهاش - معجبتنيش - الا	3
محدش - مقدرتش	2
مبتستاهل	2
مكنتش - مكنش - ماكنتش - معجبنيش - معجبتنيش - مفهمتش - مفهمتهاش - مبقتش - مابقتش - معرفش - معرفتش - ميستحقش - ميخلكش - ملوش - مفيهوش - ماينفعش - متتبعتش - معتش - مبقتش - مبيضحكش - منبسطش - محستهاش - متعميلهاش - بلاش - عديم - معاد - مب	1

TABLE IV. THE COMMON SPECULATION CUES IN NSAR

Normalized Speculation Cues	Frequency
لو	41
اعتقد	11
كانت	8
او	6
قد	6
اظن	5
ممكن	5
ربما	4
يمكن - لااعتقد - معظم - احيانا	3
اتمنى - ولا - تقريبا - لاادرى - إذا	2
مما يثيرالشك - بالرغم من الشكوك - فعلا كان - نوعا ما - لا اظن - ما بين - غالبا - ياريت - تبدو - تاكد - ان - لما	1

Table V displays the average, minimum, and maximum scope lengths for both negation and speculation for each topic. For the negation scope, the minimum and average scope lengths are nearly identical, but there is a notable variation in the maximum scope length for each topic. This notice in books and software topics usually negate the longest part of the sentence. Table V also shows that the speculated words within a sentence are longer than the negated words because the speculation structures usually affect the whole sentence, as described in the annotation guidelines.

TABLE V. NSAR NEGATION AND SPECULATION SCOPE LENGTH

Corpus	Topic	Negation Scope			Speculation Scope		
		Max	***Min***	***Avg***	***Max***	***Min***	***Avg***
LABR	Books	66	2	22	82	2	32
LAMR	Touristic Attractions	45	3	21	13	13	13
	Hotels	33	7	17	23	21	22
	Products	56	3	22	65	12	35
	Resturants 1	50	2	17	86	10	35
	Restaurant 2	44	3	26	48	10	27
MASC	Software	60	2	20	62	8	32
	Products	40	5	30	52	52	52
All		**66**	**2**	**21**	**86**	**2**	**31**

Table VI presents the distribution of negated and speculated sentences based on the overall polarity of the sentence. On average, the number of sentences with negation structures and positive polarity is the same as negative polarity. Nonetheless, the number of negation cases in the software topic with negative polarity is more than the cases with positive polarity. In addition, the speculative contents within positive polarity account for 66% of the corpus speculation cases as it is the majority in the books and software topics. According to our observation, the book's topic includes most negation and speculation cases, which are typically used to cancel something negative about the books. Furthermore, most of the software advantages or features are negated or speculated.

Fig. 1 and Fig. 2 demonstrate the number of negation cases in each sentence within the three sub-corpora. The number of negated sentences that include more than two negation scopes in one sentence is 173, accounting for 20% of the negation cases in the NSAR corpus. However, there are only three sentences with two speculation scopes. This finding further proves that the speculative content in the review domain includes the entire sentence as long as the polarity.

TABLE VI. NEGATION AND SPECULATION SENTENCES PER POLARITY

Corpus	Topic	Negation		Speculation	
		Pos	***Neg***	***Pos***	***Neg***
LABR	Books	165	83	29	17
LAMR	Touristic Attractions	17	3	1	0
	Hotels	7	0	2	0
	Products	114	170	16	14
	Resturants 1	59	39	7	3
	Restaurants 2	13	20	2	1
MASC	Software	54	112	23	5
	Products	1	5	0	1
Total		**430**	**432**	**80**	**41**

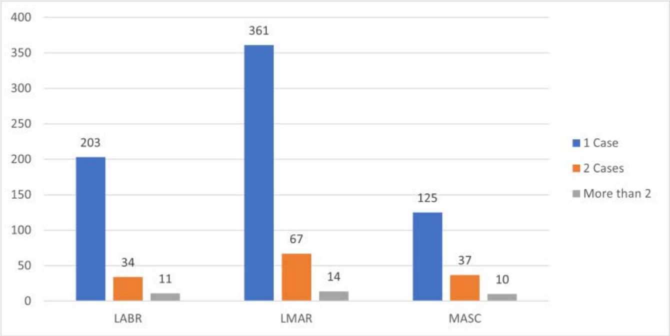


Fig. 1. The Distribution of Negation Structures per Sentence.

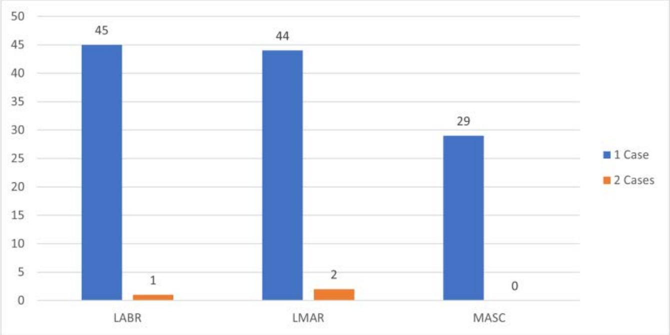


Fig. 2. The Distribution of Speculation Structures per Sentence.

VI. Conclusion and Future Work

The DA texts are used in people's day-to-day conversations on social media platforms and review websites. Many research groups worked on the sentiment analysis task, and some of them considered the negation linguistic feature and highlighted its significance using simple rules. However, researchers still have challenges in addressing various structures of the negation phenomenon as long as the speculation. This paper presented the first Arabic corpus in the review domain annotated with negation and speculation (NSAR) to tackle these challenges using supervised learning techniques. This corpus was annotated by two Arabic native speakers who adhered to strict annotation guidelines that were reviewed by a linguist expert. The Cohen's Kappa coefficients were used to measure annotator agreement and obtained 95 and 80 for negation and speculation, respectively. The results show that the annotation guidelines were written clearly. NSAR will be made available, which will contribute to the detection of negation and speculation, as well as the sentiment analysis task. The future work includes extending the corpus by annotating the events element as long as the negation focus. In addition, we plan to apply the recent deep learning techniques on this corpus to study the impact of negation and speculation on various ANLP tasks.

References

[1] Velldal, L. Øvrelid, J. Read, and S. Oepen, “Speculation and Negation: Rules, Rankers, and the Role of Syntax,” Computational Linguistics, vol. 38, no. 2, pp. 369–410, Jun. 2012.

[2] J. H. Greenberg, “Universals of human language,” Stanford University Press, vol. 4, 1978.

[3] S. Kaddoura, M. Itani, and C. Roast, “Analyzing the Effect of Negation in Sentiment Polarity of Facebook Dialectal Arabic Text,” Appl. Sci., vol. 11, no. 11, p. 4768, May 2021.

[4] O. Solarte Pabón, M. Torrente, M. Provencio, A. Rodríguez-Gonzalez, and E. Menasalvas, “Integrating Speculation Detection and Deep Learning to Extract Lung Cancer Diagnosis from Clinical Notes,” Appl. Sci., vol. 11, no. 2, p. 865, Jan. 2021.

[5] C. Dalloux et al., “Supervised learning for the detection of negation and of its scope in French and Brazilian Portuguese biomedical corpora,” Natural Language Engineering, vol. 27, no. 2, pp. 181–201, Mar. 2021.

[6] V. Vincze, G. Szarvas, R. Farkas, G. Móra, and J. Csirik, “The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes.,” BMC bioinformatics, vol. 9, no. 11, p. Suppl 11-S9, Nov. 2008.

[7] A. Mahany, H. Khaled, N. S. Elmitwally, N. Aljohani, and S. Ghoniemy, “Negation and Speculation in NLP: A Survey, Corpora, Methods, and Applications,” Applied Sciences, vol. 12, no. 10, p. 5209, May 2022.

[8] S. M. Jiménez-Zafra, R. Morante, E. Blanco, M. T. Martín-Valdivia, and L. Alfonso Ureña-López, “Detecting negation cues and scopes in Spanish,” in LREC 2020 - 12th International Conference on Language Resources and Evaluation, 2020, pp. 6902–6911.

[9] M. M. Hossain, A. Anastasopoulos, E. Blanco, and A. Palmer, “It's not a Non-Issue: Negation as a Source of Error in Machine Translation,” in Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 3869–3885.

[10] W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan, “A simple algorithm for identifying negated findings and diseases in discharge summaries,” Journal of Biomedical Informatics, vol. 34, no. 5, pp. 301–310, Oct. 2001.

[11] H. Fei, Y. Ren, and D. Ji, “Negation and speculation scope detection using recursive neural conditional random fields,” Neurocomputing, vol. 374, pp. 22–29, Jan. 2020.

[12] A. Mahany et al., “Supervised Learning for Negation Scope Detection in Arabic Texts,” in Tenth International Conference on Intelligent Computing and Information Systems (ICICIS), 2021, pp. 177–182.

[13] A. Alsayat and N. Elmitwally, “A comprehensive study for Arabic Sentiment Analysis (Challenges and Applications),” Egyptian Informatics Journal, vol. 21, no. 1, pp. 7–12, Mar. 2020.

[14] N. Y. Habash, Introduction to Arabic natural language processing, 1st ed., vol. 3, no. 1. Morgan and Claypool Publishers, 2010.

[15] A. Elnagar, S. M. Yagi, A. B. Nassif, I. Shahin, and S. A. Salloum, “Systematic Literature Review of Dialectal Arabic: Identification and Detection,” IEEE Access, vol. 9, pp. 31010–31042, 2021.

[16] S. R. El-Beltagy, “NileULex: A phrase and word level sentiment lexicon for Egyptian and modern standard Arabic,” Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016, pp. 2900–2905, 2016.

[17] N. P. Cruz, M. Taboada, and R. Mitkov, “A machine-learning approach to negation and speculation detection for sentiment analysis,” Journal of the Association for Information Science and Technology, vol. 67, no. 9, pp. 2118–2136, Sep. 2016.

[18] A. E.-D. Hamouda and F. E. El-taher, “Sentiment Analyzer for Arabic Comments System,” International Journal of Advanced Computer Science and Applications, vol. 4, no. 3, pp. 99–103, 2013.

[19] R. M. Duwairi and M. A. Alshboul, “Negation-Aware Framework for Sentiment Analysis in Arabic Reviews,” in 2015 3rd International Conference on Future Internet of Things and Cloud, 2015, pp. 731–735.

[20] N. El-Naggar, Y. El-Sonbaty, and M. A. El-Nasr, “Sentiment analysis of modern standard Arabic and Egyptian dialectal Arabic tweets,” in 2017 Computing Conference, 2017, pp. 880–887.

[21] A. Assiri, A. Emam, and H. Al-Dossari, “Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis,” Journal of Information Science, vol. 44, no. 2, pp. 184–202, Jan. 2018.

[22] N. Alalyani and S. Larabi, “NADA: New Arabic Dataset for Text Classification,” International Journal of Advanced Computer Science and Applications, vol. 9, no. 9, pp. 206–212, 2018.

[23] M. Aly and A. Atiya, “LABR: A large scale arabic book reviews dataset,” in ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, 2013.

[24] H. ElSahar and S. R. El-Beltagy, "Building Large Arabic Multi-domain Resources for Sentiment Analysis," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9042, pp. 23–34, 2015.

[25] T. Al-Moslmi, M. Albared, A. Al-Shabi, N. Omar, and S. Abdullah, "Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis," Journal of Information Science, vol. 44, no. 3, pp. 345–362, Jun. 2018.

[26] M. Neves and J. Ševa, "An extensive review of tools for manual annotation of documents," Briefings in Bioinformatics, vol. 22, no. 1, pp. 146–163, Jan. 2021.

[27] R. Eckart de Castilho et al., "A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures," in workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH), 2016, pp. 76–84.

[28] J. Cohen, "A Coefficient of Agreement for Nominal Scales," Educational and Psychological Measurement, vol. 20, no. 1, pp. 37–46, Apr. 1960.

[29] N. Pröllochs, S. Feuerriegel, B. Lutz, and D. Neumann, "Negation scope detection for sentiment analysis: A reinforcement learning framework for replicating human interpretations," Information Sciences, vol. 536, pp. 205–221, Oct. 2020.

APPENDIX I

TABLE VII. ANNOTATION GUIDELINES EXAMPLES

No.	Arabic Text	Transliteration*	English Translation
1	أحلام اجمل الايام كانت هناك و ذكريات لا تنسى مكان رائع يجمع الكل فى حركه و ضحك وحياه	Aḥlām ajmal al-Ayyām kānat hunāk wa Dhikrayāt lā tunsá makān rā'i' yajma'u al-Kull fī ḥrkh wa ḍaḥika wḥyāh	The most beautiful days were spent there with unforgettable memories; it is a wonderful place that brings everyone together in liveliness, laughter and life.
2	لو كان سعره اقل وفيه فلاش كنت قيمته اكار من كدا فشل واضح من سوني انها ماحطتش ليه فلاش	Law kāna si'ruhu aqall wa-fīhi Flāsh Kunt qymth akār min kdā fashal Wāḍiḥ min Sūnī annahā māḥṭtsh Līh Flāsh	If it had a lower price and had a flash, it would have been worth a lot more than this. It is a clear failure from Sony that they didn't add a flash.
3	جهاز جبار ويتفوق علي نظرائة من الايباد والسامسونج بجد رابع بس لم ياخذ حجم الدعايه المطلوبه	Jihāz Jabbār wytfwq 'Alī nẓrā'h min alāybād wālsāmswnj bi-jadd rāá' Bass lam yākhdh ḥajm ald'āyh almṭlwbh	This an excellent device that outperforms its counterparts from iPad and Samsung which is really great, but it did not receive enough publicity.
4	7 روايات في رواية ، تهكم وسخرية وأدب وأشياء اخرى لن تمل منها	7 Riwāyāt fī riwāyah, thkm wskhryh wa-adab wa-ashyā' ukhrá lan tml minhā	This book has seven novels in a novel: sarcasm, irony, literature and other things that you will not get bored of.
5	يستاهل لانه دعم اللغة العربية لكن ياريت يثبت البرنامج على رقم المستخدم وليس على حساب	Ystāhl lānh Da'm al-lughah al-'Arabīyah lākin yāryt yuthbatu al-Barnāmaj 'alá raqm al-mustakhdam wa-laysa 'alá ḥisāb	It is worth it because it supports the Arabic language, but I hope that I can register with my mobile number not my account number.
6	الذي فاجئني وزعجني هو عدم وجود ريموت فيها بالمقارنة مع السعر	Alladhī fāj'ny wz'jny huwa 'adam wujūd rymwt fīhā bi-al-muqāranah ma'a al-si'r	What shocked and annoyed me was the lack of a remote control in it compared to the price.
7	تطبيق رائع وفيه خصوصيه محدش يعرف دخلت امتي خرجت امتي انت حتي لو بتكتب اللي قدامك مش بيعرف	Taṭbīq rā'i' wa-fīhi khṣwṣyh Maḥaddish ya'rifu dkhlt amty kharajat amty anta ḥattá Law btktb Illī qdāmk mish by'rf	This is a wonderful application that protects your privacy; no one knows when you logged in or logged out. Even while typing, the person in front of you will not know that you are typing.
8	فكرني بفلاش و سماش والحاجات دي بس مفيش كلمات متقاطعة	Fkrny bflāsh wa smāsh wālḥājāt Dī Bass mfysh Kalimāt mutaqāṭi'ah	It reminded me of 'Flash and Smash' and these things, but this one does not have crossword puzzles.
9	مررررره حلو وعصيراته فرش ومو حاطين له لاسكر ولا مويه كله فرش	Mrrrrrh Ḥulw w'ṣyrāth farsh wmw ḥāṭyn la-hu lāskr wa-lā mwyh kullahu farsh	It is extremely delicious, and its juices are fresh, and they do not add sugar or water; it is all fresh.
10	نصيحتي ان في محلات فطاير تانية في حدايق حلوان ممكن تكون أفضل كتير	Naṣīḥatī an fī maḥallāt fṭāyr tānyh fī Ḥadāīq Ḥulwān mumkin takūn afḍal kitīr	My advice is that there are other pastry shops in Hadyaa Helwan that could be much better.
11	اول روايه اقرأها لباولو كويلهو ولااعتقد انها الاخيره	Awwal riwāyah aqr'hā lbāwlw kwylhw wlāā'tqd annahā alākhyrh	This is the first novel I read for Paulo Coelho, and I don't think it will be the last.
12	لم استسيغ نزار فى شعر الفصحى اعتقدوا فى الشعر الحر اكثر ابداعا	Lam astsygh Nizār fī shi'r al-fuṣḥá a'tqdwā fī al-shi'r al-Ḥurr akthar abdā'ā	I did not like Nizar in classical poetry. I think in free verse he is more creative.
13	فندق مريح صراحة اسعارة جيدة ما بين 150 الى 300 ريال الليلة ونظيف جدا	Funduq mryḥ ṣrāḥh as'ārh Jīdah mā bayna 150 ilá 300 Riyāl al-laylah wa-Naẓīf jiddan	It is a comfortable hotel, frankly; it has good prices, between 150 to 300 riyals per night, and it is very clean.
14	أنصح بيه أي حد بيدرس هندسة كهربية أو عايز يدرسها هتوفر عليه جنون كتير	Anṣḥ Bīh Ayy ḥadd bydrs Handasat khrbyh aw 'āyiz ydrshā htwfr 'alayhi Junūn kitīr	I recommend it to anyone who studies electrical engineering or wants to study it; it will save a lot for him.
15	ربما يكون الكتاب جيداً ولكن بروز شخصية الكاتب المتملقه تفسد ذلك؟	Rubbamā yakūn al-Kitāb jayyidan wa-lakin Burūz shakhṣīyah al-Kātib almtmlqh tufsidu dhālika?	The book may be good, but the author's fawning character spoils it.
16	ما إن رأيت ولا سمعت بمثله	Mā Inna ra'aytu wa-lā sami't bi-mithlih	I have neither seen nor heard of anything like it.
17	ما هذه الرومانسية الحالمة وما هذا الاسلوب الناعم الجميل هذه الرواية من اجمل ما قرات على الاطلاق	Mā Hādhihi al-rūmānsīyah al-ḥālimah wa-mā Hādhā al-uslūb al-Nā'im al-jamīl Hādhihi al-riwāyah min ajmal mā qrāt 'alá al-iṭlāq	What is this dreamy romance, and what is this soft and beautiful style? This novel is one of the most beautiful novels I have ever read.
18	أليس هذا بالحق	Alīs Hādhā bi-al-Ḥaqq	Isn't that right?
19	مَا هَٰذَا بَشَرًا إِنْ هَٰذَا إِلَّا مَلَكٌ كَرِيمٌ	Mā haādhā basharan in haādhā illā malakun karīm	This is not a man; this is none but a noble angel **
20	جدة غير	Jiddah ghayr	Jeddah is different/unique.

21	كتاب غير مجرى تفكيرى خلاه اوسع خلانى أثق اوووى فى العلامات	Kitāb ghayr majrá tfkyrá khlāh awsaʻ khlāná athq awwwá fī al-ʻalāmāt	This book changed my way of thinking; it broadened my mind and made me trust the signs strongly.
22	تجننننن انا بصراحه مش بس الفندق جميل كل حاجه زورتها كانت جميله قوي قوي قوي قوي	Tjnnnnnn anā bṣrāḥh mish Bass al-Funduq Jamīl kull ḥājh zwrthā kānat Jamīlah Qawī Qawī Qawī Qawī	Amazing! Not only is the hotel beautiful but also everything I visited there was very very very very beautiful.
23	أجمل ما فيه هو إفطاره	Ajmal mā fīhi huwa ifṭārh	The best thing about it is its breakfast.
24	كتاب خفيف و واقعي و بعيد عن المبالغة تماما كل شيء فيه حقيقي	Kitāb khafīf wa wāqiʻī wa baʻīd ʻan al-Mubālaghah tamāman kull Shay’ fīhi ḥaqīqī	This is a light and realistic book which is absolutely far from exaggeration; everything in it is real.
25	وَتَرَى النَّاسَ سُكَارَىٰ وَمَا هُم بِسُكَارَىٰ	Watará alnnāsa sukāráā wamā hum bisukāráā	You will see the people [appearing] intoxicated while they are not intoxicated [**]
26	الكلمة قد تفعل فى الانسان ما لم تفعلة الادوية القوية لك كل قدير	al-Kalimah qad tfʻl fī al-insān mā lam tfʻlh al-adwīyah al-qawīyah laka kull qdyr	The effect of a word may be stronger than the effect of medicines.
27	السلعة ليست بالجودة المطلوبة وقد اشتخدمتها لمرة واحدة فقط ولم ارجع لاستخدامها مرة ثانية	Alslʻh laysat bāljwdh al-maṭlūbah wa-qad ashtkhdmthā li-marrah wāḥidah faqaṭ wa-lam arjʻ lāstkhdāmhā marrah thānīyah	The item is not of the expected quality, and I only used it once and did not use it again.
28	جلسات رائعة جلسات المطعم الخارجية رائعة خصوصا في فصل الربيع والشتاء اما الاكل فجيد نوعا ما	Jalasāt rā’iʻah jalasāt almṭʻm al-khārijīyah rā’iʻah khṣwṣā fī Faṣl al-Rabīʻ wa-al-shitā’ amā alākl fjyd nwʻā mā	The atmosphere of the outdoor restaurant is wonderful, especially during spring and winter, but the food is not that good.
29	وَجَعَلَ لَكُمُ السَّمْعَ وَالْأَبْصَارَ وَالْأَفْئِدَةَ ۙ لَعَلَّكُمْ تَشْكُرُونَ	Wajaʻala lakumu alssamʻa wāl’abṣāra wāl’af’idata ۙ laʻallakum tashkurūn	He [Allah] made for you hearing and vision and intellect that perhaps you would be grateful [**]
30	الَّذِينَ يَظُنُّونَ أَنَّهُم مُّلَاقُو رَبِّهِمْ وَأَنَّهُمْ إِلَيْهِ رَاجِعُونَ	Alladhīna yaẓunnūna annahum mmulāqū rabbihim wa’annahum ilayhi rājiʻūn	Who are certain that they will meet their Lord and that they will return to Him [**]

[*] The transliteration is accomplished by the developed tool at CAMeL Lab, New York Abu Dhabi University (http://romanize-arabic.camel-lab.com/)

[**] The source of translation is King Saud University Mushaf (https://quran.ksu.edu.sa/)

Drought Prediction and Validation for Desert Region using Machine Learning Methods

Azmat Raja, Gopikrishnan T
Department of Civil Engineering
National Institute of Technology, Patna, India

***Abstract*—Drought prediction serves as an early warning to the effective management of water resources to avoid the drought impact. The drought prediction is carried out for arid, semi-arid, sub-humid, and humid climate types in the desert region. The drought is predicted using Standardized precipitation evapotranspiration index (SPEI). The application of machine learning methods such as artificial neural network (ANN), K-Nearest Neighbors (KNN), and Deep Neural Network (DNN) for the drought prediction suitability is analyzed. The SPEI is predicted using the aforesaid machine learning methods with inputs used to calculate SPEI. The predictions are assessed using statistical indicators. The coefficient of determination of ANN, KNN, and DNN are 0.93, 0.83, and 0.91 respectively. The mean square error of ANN, KNN, and DNN are 0.065, 0.512, and 0.52 respectively. The mean absolute error of ANN, KNN, and DNN are 0.001, 0.512, and 0.01 respectively. Based on results of statistical indicator and validations it is found that DNN is suitable to predict drought in all the four types of desert region.**

Keywords—Drought; SPEI; machine learning; water resources; prediction

I. Introduction

Drought is a recurring natural phenomenon characterized by prolonged water crisis driven by below-normal precipitation over a considerable length of time ranging from months to years [11]. Because of its complex nature and extensive occurrence, it is difficult to describe drought and identify its characteristics [16]. For sustainable agricultural activity and water resource management, accurate drought prediction and management are critical. The duration, frequency, intensity, and geographic distribution of rainfall, as well as the water needs of humans, animals, crops, and the region's vegetative cover, assess the severity of the drought. Drought is divided into four categories based on its effects: meteorological drought, agricultural drought, hydrological drought, and socioeconomic drought [8] [25]. Drought can reduce food production and have an impact on a community's socioeconomic viability [32].

Agricultural drought analysis and forecasting are more important than other types of drought in India since it is an agrarian country with 68 percent of the population dependent on agriculture. Due to a lack of water, anomalous rainfall, and harsh climatic conditions, the dry region of western India is at risk of severe droughts. The state of Rajasthan in India comprises of desert region which is classified as arid, semi-arid, sub-humid, and humid. Rajasthan is vulnerable to drought due to harsh climate. Drought is unavoidable, but it can be minimized by careful planning and preparation. Effective drought responses do not come from simply understanding the situation. While resources should be invested in enhancing the accuracy of drought monitoring and early warning systems, equal or more emphasis should be placed on upgrading drought governance structures to ensure a more effective division of responsibilities between national and local governments. As a result, numerous studies have been conducted in order to provide an objective and quantitative assessment of drought severity. Drought indices, which are proxies based on climate data and assumed to appropriately characterize the degree of drought hazard, are commonly used to quantify the effects of drought.

Drought indices have been developed in large numbers and are widely used in drought evaluation, monitoring, and forecasting. The standardized precipitation evapotranspiration index (SPEI) [36] is a developed form of standardized precipitation index (SPI) [27] by taking monthly climatic water balance. The SPEI's multi-scalar properties allow it to distinguish between different drought types and impacts, which is a significant advantage over the most generally used drought indices, which assess the effect of potential evapotranspiration on drought severity [5].

Machine learning is the subset of artificial intelligence. Machine learning (ML) algorithms are a set of commands that allow systems to learn and improve from prior data without requiring complex programming. ML techniques have been used to implement prediction or forecasting of drought. These algorithms work by simulating a model from input datasets known test sets, and then using the model findings to forecast, predict, or make various types of judgments in various application domains. K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), and Deep Neural Network (DNN) machine learning techniques are extensively used in prediction of drought [1]-[4], [7], [9]-[10], [12]-[15], [17]-[24], [28]-[31], [33]-[35].

KNN is also called Lazy Learner. It does not require any training. It stores the training dataset and only uses it to make real-time predictions. As a result, the KNN method is much faster than other algorithms that require training. KNN is a simple algorithm to use. KNN can be implemented with only two parameters: the value of K and the distance function. ANN have a number of advantages, including the ability to detect complex nonlinear relationships between dependent and independent variables, the ability to recognize all potential relationships between predictor factors [10]. ANN is suitable for both large and small data sets. Deep learning is significantly responsible for the current boom in artificial intelligence (AI) usage. DNN has the biggest advantage in that it learns high-level features from data in a gradual manner. DNN gives the high quality results over other traditional methods. Overfitting is a major problem in neural networks which can be easily handled in DNN using dropout layer.

Various studies have used machine learning approaches to predict the drought using SPEI. For drought estimation, we take into account all climatic parameters that influence evapotranspiration. All of the activation functions are used to determine the best activation function for drought prediction. In this study we use SPEI drought indices to measure drought in western Rajasthan from 1979 to 2013. Drought prediction and selection of acceptable methodologies for drought measures in dry regions utilizing ANN, KNN, and Deep Neural Network machine learning algorithms.

II. Materials and Methods

The computation process of the drought prediction using both SPEI drought indices and machine learning techniques is shown in work flow diagram (Fig. 1). The working procedure start from the data collection and end at accuracy assessments of machine learning models.

A. Study Area

Rajasthan is located in north-western India between latitudes of 230 N and 300 N and longitudes of 690 E and 780 E. Fig. 2 depicts the study area in western Rajasthan, which consists of 12 districts in the state of Rajasthan. Due to insufficient rainfall and vast desert, the 12 districts are at risk of drought. Jaisalmer and Bikaner are in the arid region; Ganganagar, Hanumangarh, and Jodhpur are in the semi-arid region; Jalor, Jhunjhunu, Pali, and Sikar are in the sub-humid region; and Barmer district is in the humid region.

B. Data Collection

Daily and monthly weather data for all 12 districts in western Rajasthan for 24 weather stations were retrieved from global weather data (https://globalweather.tamu.edu/) from 1979 to 2013. Maximum and minimum temperatures, precipitation, wind speed, humidity, and solar radiation are all included in the weather data.

C. Standardized Precipitation Evapotranspiration Index (SPEI)

Vicente et al. [36] created SPEI, which is derived from SPI and uses both precipitation and potential evapotranspiration (PET) as input parameters to calculate the monthly climatic water balance.

$$D = P - PET \qquad (1)$$

Where D is a simple aggregation of water shortages or excess over a period of time, P is the precipitation in mm, and PET is potential evapotranspiration in mm. The PET is computed using penman- Monteith (PM) equation using meteorological data [5-6].

The SPEI value can be calculated using the standardized values of F(x) [36].

$$SPEI = W - \frac{C_0 + C_1W + C_2W^2}{1 + d_1W + d_2W^2 + d_3W^3} \qquad (2)$$

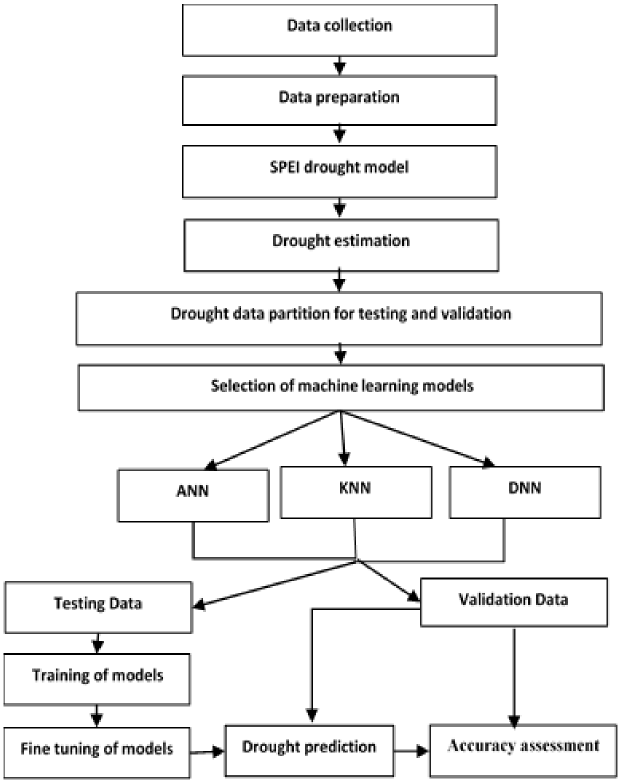


Fig. 1. The Workflow Diagram.

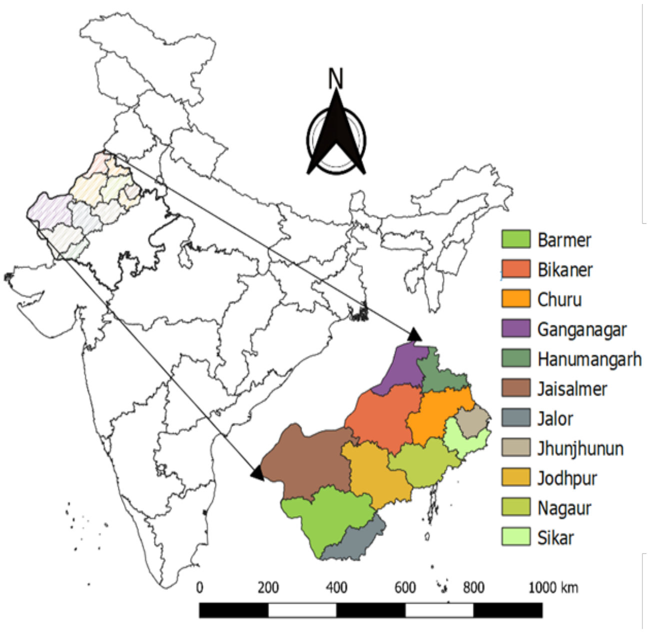


Fig. 2. The Map of the Study Area.

Where, $W = \sqrt{-2lnp}$ for P ≤ 0.5 and P is the probability of exceeding a determined D value, P =1-F(x). If P > 0.5, then P is replaced by 1 - P and the sign of the resultant SPEI is reversed. The constants are C_0= 2.515517, C_1 = 0.802853, C_2= 0.010328, d_1 = 1.432788, d_2 = 0.189269 and d_3= 0.001308.

According to the classifications of [27] [5] the SPEI can be classified mainly in five classes as shown in Table I.

TABLE I. SPEI DROUGHT CATEGORY CLASSIFICATION

Drought class	SPEI
No-Drought	Greater than -0.5
Mild	-0.5 to -0.99
Moderate	-1 to -1.49
Severe	-1.50 to -1.99
Extreme	Less than -2

D. Machine Learning Methods

Artificial neural network (ANN): McCulloch and Pitts [26] created the technique in the 1940s, and it has since evolved alongside developments in calibrating methodologies. One of the benefits of the ANN method is that the simulation is not required to completely characterize the intermediate relationships [10], [12], [21]. An ANN consists of a series of input layers, hidden layers, and output layers, each with its unique weightage as shown in Fig. 3. In this study we used 7 inputs variables and two hidden layers (first hidden layer consists eight nodes and second hidden layer consists five nodes) to compute the SPEI at output node.

1) K-Nearest Neighbors (KNN): The K-Nearest Neighbors (KNN) algorithm is a supervised machine learning algorithm that can be used to address classification and regression problems. The KNN is successfully applied in drought forecasting globally [13]-[14] [17] [28]. The basic logic behind the KNN is computing the Euclidean distance, the other alternative distance can be used are Manhattan distance, Hamming Distance, and Minkowski distance. The outcome of a KNN regression is the mean of the k closest data points. We choose odd numbers as k as a rule of thumb. KNN is a lazy learning model in which only runtime computations are performed. KNN has advantage over ANN as neural networks need large training data. The importance of meteorological data and PET on drought prediction is shown in Fig. 4.

2) Deep Neural Network (DNN): Deep learning has risen in popularity in recent years as a result of its superiority in prediction when compared to traditional machine learning approaches. Deep learning takes a lot of data because the network learns on its own. Traditional machine learning is just a set of algorithms for parsing and learning from data. The DNN has been successfully used for prediction in the past with positive results [7], [18], [21], [23], [33]-[34]. We used Keras and tensor flow package in R which is based on deep learning. The high-level interface and the automatic differentiation feature of TensorFlow make simple to implement the algorithm in an efficient manner [24]. The process of computation involved in prediction is shown in Fig. 5.

Relu, Selu, Tanh, and Sigmoid are the most common activation functions utilized in DNN. As stated in Table II, we used all four activation functions as well as their combinations. The DNN has a high prospect of overfitting the model in most cases. We used 20% of the training data for validation, as indicated in Fig. 6, to avoid overfitting the model.

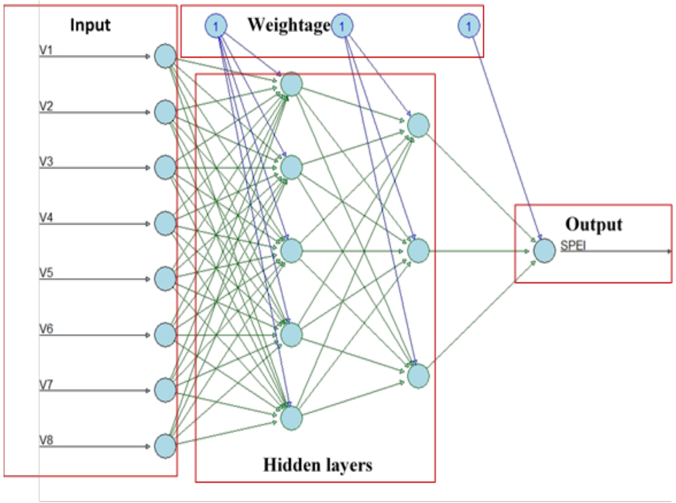


Fig. 3. The Components of ANN.

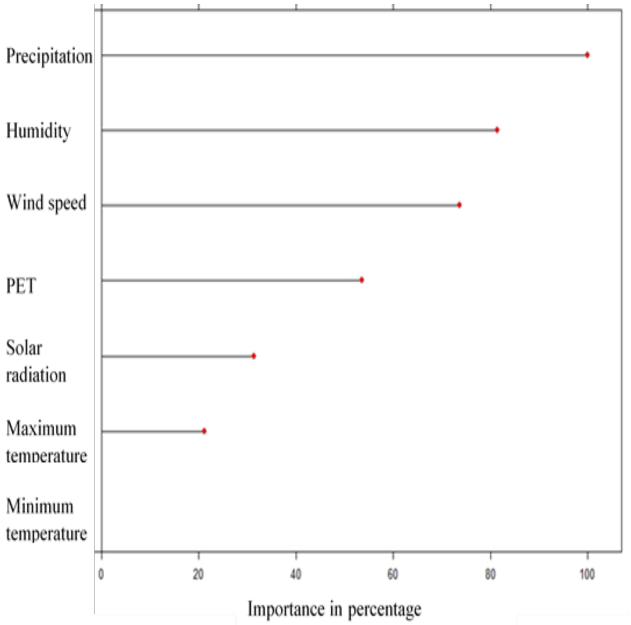


Fig. 4. The Importance of Meteorological Parameters on Drought Prediction using KNN.

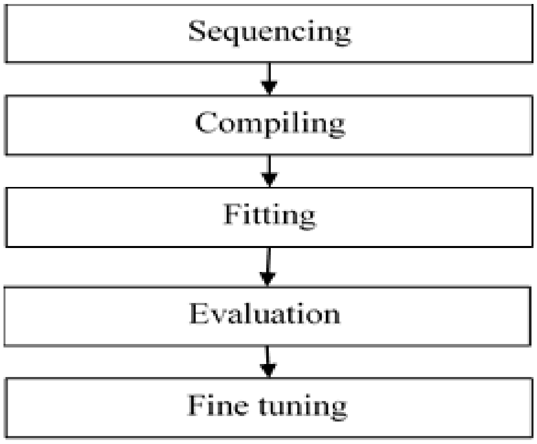


Fig. 5. Computation Process of Drought.

TABLE II. ACTIVATION FUNCTION USED IN DNN MODELLING

Activation function	Abbreviation
Relu without fine-tuning	M
Relu	M1
Selu	M2
Tanh	M3
Sigmoid	M4
Combination of Relu, Selu, and Tanh	M5
Combination of Relu, Selu, Sigmoid, and Tanh	M6

3) Performance assessment: The mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) are three typical performance indicators used to assess the models performance.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |Actual - Predicted| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Actual - Predicted)^2} \quad (4)$$

$$R^2 = 1 - \frac{Squared\ sum\ error\ of\ regrssion\ line}{Squared\ sum\ error\ of\ mean\ line} \quad (5)$$

Where, N is the number of samples.

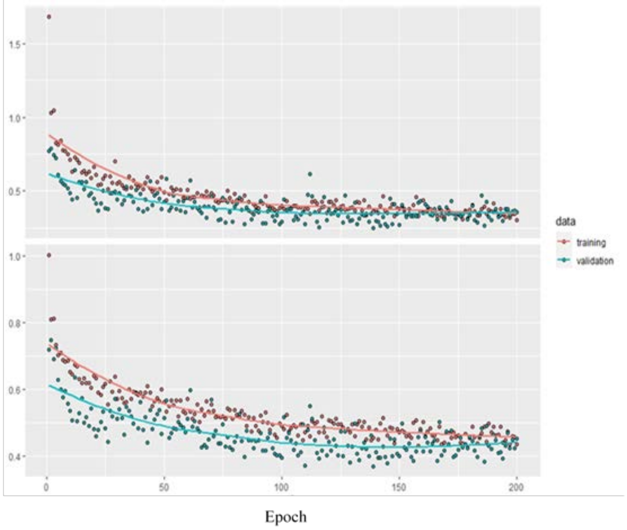


Fig. 6. Validation Spilt to avoid Overfitting of the Model.

III. RESULT AND DISCUSION

A. Temporal Variation of SPEI

The temporal variation of SPEI from 1979 to 2013 for different regions of western Rajasthan is shown in Fig. 7. The variation of drought pattern is varying with climatic condition. Each region observed severe to extreme drought in the year 1987, 1999, 2002, 2006, and 2009. The arid and semi-arid region showed more vulnerable to drought than the humid and sub-humid region Due to their location in the driest section of the Rajasthan, the districts of Jaisalmer and Bikaner experienced severe drought whereas other districts had moderate drought. Drought is less of a concern in Barmer, but it is more of a problem in Jaisalmer.

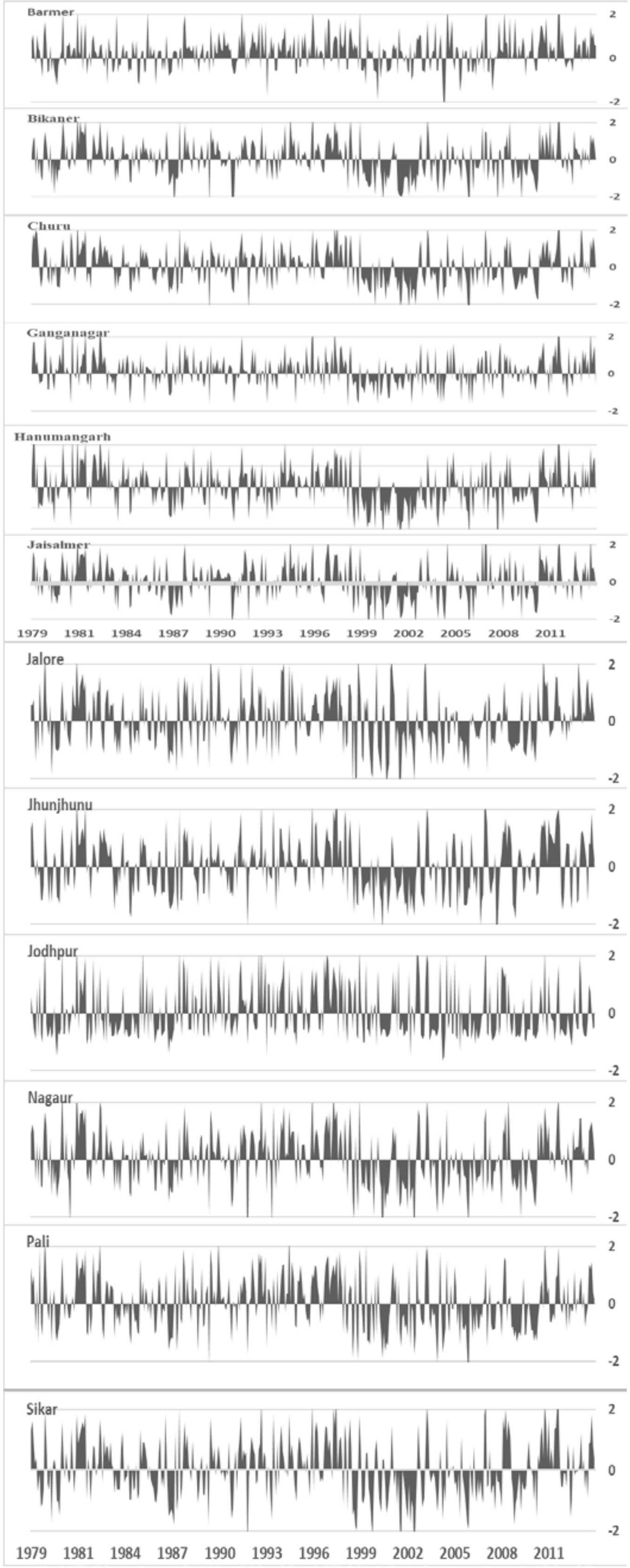


Fig. 7. Temporal Variations of SPEI over the Years 1979-2013.

B. Validation of ANN Drought

The variation of drought predicted by ANN for four different meteorological conditions is shown in Fig. 8. With slightly high or low SPEI values, the predicted values showed a similar pattern of variation as the test target. All of the regions showed good results, although the humid zone produced slightly better results. We used normalized data because there is always a problem with convergence with ANN when dealing with large data sets.

C. Validation of KNN Drought

The variation SPEI values predicted by KNN is shown Fig. 9. All the region showed similar kind of variation. The ANN showed slight better variation than KNN with target SPEI values. For sub-humid and humid locations, the K-NN based SPEI values best matched lower test target SPEI values, indicating that K-NN is better suited for drought prediction in humid regions.

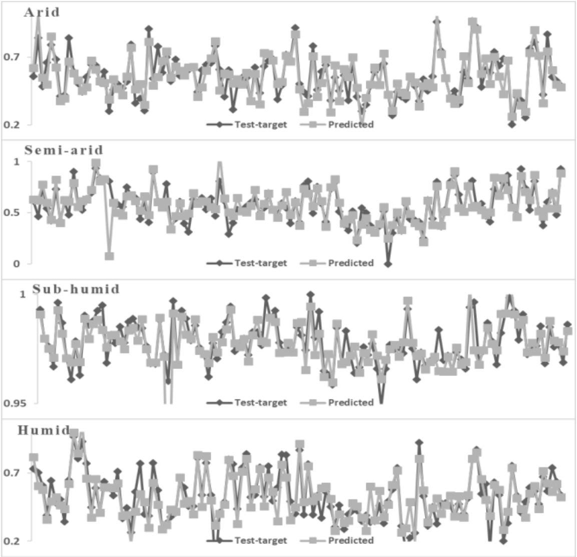


Fig. 8. Comparison of ANN Predicted and Test Target SPEI.

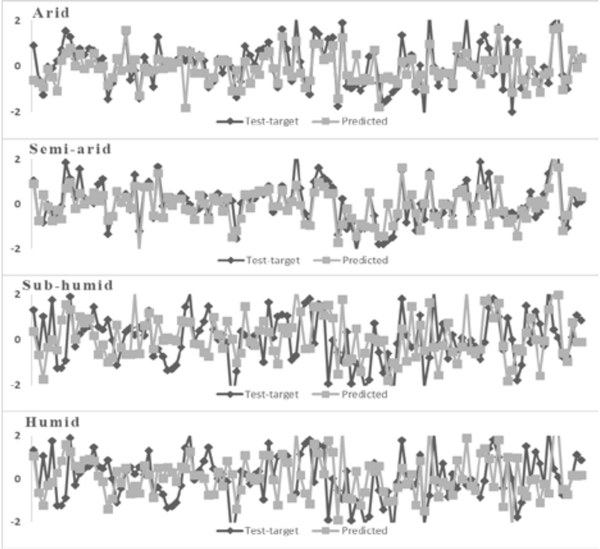


Fig. 9. Comparison of KNN Predicted and Test Target SPEI.

D. Validation of DNN Drought

The variation SPEI values predicted by DNN using a fine-tuned relu activation functions shown Fig. 10. The variation of SPEI values predicted by the DNN showed similar pattern of variation as target value with almost same SPEI values. The results of DNN is better than the both ANN and KNN for all the regions. The predicted values do not exceed the target values because the DNN has overfitting control. All the activation functions and their combinations showed good results.

E. Performance Assesment

Three different statistical indicators are used to evaluate the drought prediction models' performance: coefficient of determination, mean square error, and mean absolute error. The performance statistics of the ANN, KNN, and DNN is shown in Fig. 11, Fig. 12, and Fig. 13 respectively for arid, semi-arid, sub-humid, and humid regions. After fine-tuning, both ANN and DNN showed considerable improvements in statistical parameters, however KNN did not show any substantial improvement. In ANN and DNN, the coefficient determination is higher, whereas in KNN, it is slightly lower but still satisfactory. The highest coefficient of determination of ANN, KNN, and DNN are 0.93, 0.83, and 0.91 in semi-arid, sub-humid, and semi-arid region respectively. All the three models resulted in low MSE and very low MAE values. All of the activation functions in DNN worked well, however Relu and the combination of activation functions worked very well. The fine-tuning of the model improves the results, but it can also lead to overfitting, therefore when fine-tuning is applied we need to examine the pattern and similarity of predicted values with target values. The activation function Relu after fine-tuning showed highest agreement in arid and humid region with R^2 values 0.87 and 0.89 respectively. The highest agreement observed in semi-arid and sub-humid region by fine-tuned Selu with R^2 values 0.91 and 0.87 respectively.

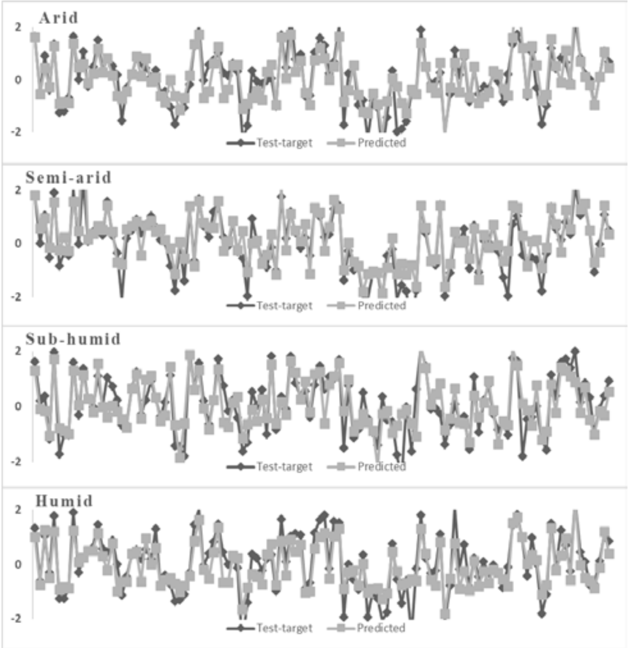


Fig. 10. Comparison of DNN Predicted and Test Target SPEI.

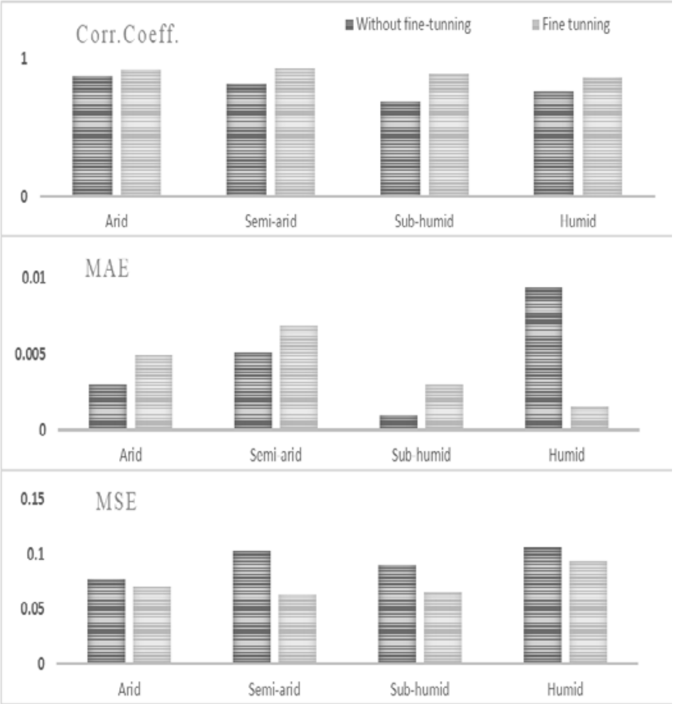


Fig. 11. Statistical Indices of ANN for different Climatic Condtions.

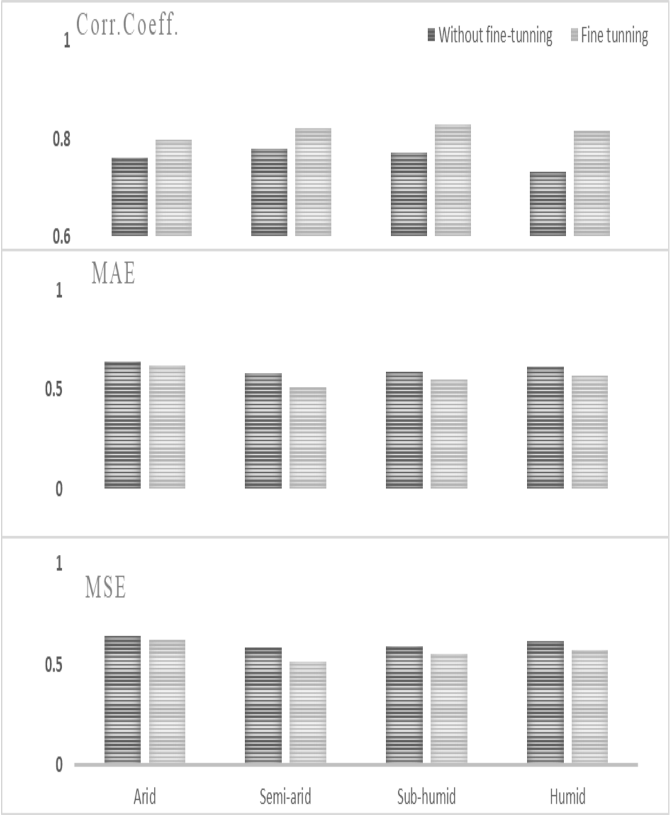


Fig. 12. Statistical Indices of KNN for different Climatic Condtions.

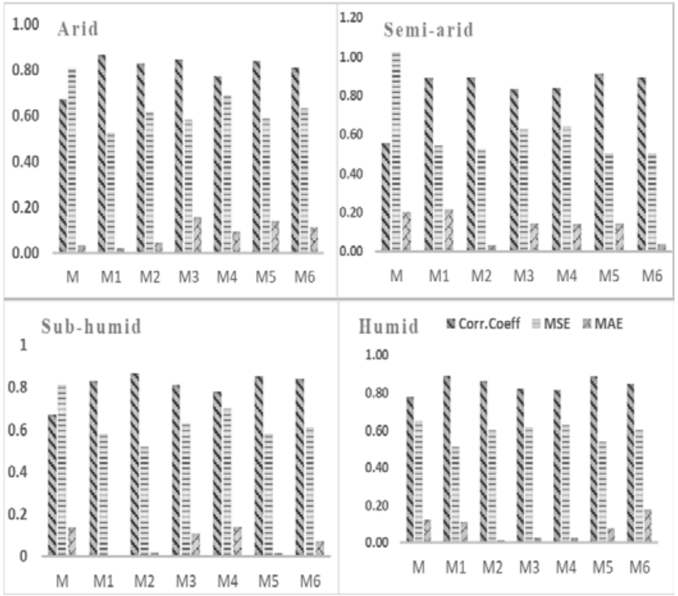


Fig. 13. Statistical Indices of DNN for different Climatic Condtions.

IV. Conclusion

The ANN, K-NN, and DNN are used to predict the drought in arid, semi-arid, sub humid, and humid conditions for the years 1979 to 2013. The KNN is one of the oldest machine learning algorithms but it is still in use for comparison to other machine learning methods. The DNN performed better than the both ANN and KNN for predicting drought in all the climate conditions. The ANN predicted better results than the KNN. It is also observed that both the ANN and KNN showed very similar results to the SPEI prediction in humid region. In all climatic conditions, DNN and ANN have larger coefficients of determination, whereas KNN has a comparatively lower coefficient of determination. The activation Relu showed best results compared to other activation function. Based on the results DNN can be used for drought prediction of any climatic condition with large data sets.

References

[1] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," arXiv preprint arXiv: 1605.07678, 2016.

[2] A. Dikshit, B. Pradhan, and A. M. Alamri, "Temporal hydrological drought index forecasting for New South Wales, Australia using machine learning approaches," Atmosphere (Basel), vol. 11, no. 6, p. 585, 2020.

[3] A. Dikshit, B. Pradhan, and A. Huete, "An improved SPEI drought forecasting approach using the long short-term memory neural network," J Environ Manage, vol. 283, p. 111979, 2021.

[4] A. Mokhtar et al., "Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms," IEEE Access, vol. 9, pp. 65503–65523, 2021.

[5] A. Raja and T. Gopikrishnan, "Drought Analysis Using the Standardized Precipitation Evapotranspiration Index (SPEI) at Different Time Scales in an Arid Region", Eng. Technol. Appl. Sci. Res., vol. 12, no. 4, pp. 9034–9037, Aug. 2022.

[6] Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56.

[7] Belayneh, A., & Adamowski, J. (2013). Drought forecasting using new machine learning methods. Journal of Water and Land Development, 18(9), 3–12. https://doi.org/10.2478/jwld-2013.

[8] Boken, V. K., Cracknell, A. P., & Heathcote, R. L. (2005). Monitoring and predicting agricultural drought: a global study. Oxford University Press.

[9] Bourquin, J., Schmidli, H., van Hoogevest, P., & Leuenberger, H. (1998). Comparison of artificial neural networks (ANN) with classical modelling techniques using different experimental designs and data from a galenical study on a solid dosage form. In European Journal of Pharmaceutical Sciences (Vol. 6).

[10] Cao, W., Wang, X., Ming, Z., & Gao, J. (2018). A review on neural networks with random weights. Neurocomputing, 275, 278–287.

[11] Dai, A. (2011). Drought under global warming: a review. Wiley Interdisciplinary Reviews: Climate Change, 2(1), 45–65.

[12] Deo, R. C., & Şahin, M. (2015). Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research, 153, 512–525. https://doi.org/10.1016/j.atmosres.2014.10.016.

[13] Fadaei-Kermani, E., Barani, G. A., & Ghaeini-Hessaroeyeh, M. (2017). Drought monitoring and prediction using K-nearest neighbor algorithm. Journal of AI and Data Mining, 5(2), 319–325.

[14] Fathabadi, A., Gholami, H., Salajeghe, A., Azanivand, H., & Khosravi, H. (2009). Drought forecasting using neural network and stochastic models. Advances in Natural and Applied Sciences, 3(2), 137–147.

[15] Fung, K. F., Huang, Y. F., Koo, C. H., & Mirzaei, M. (2020). Improved SVR machine learning models for agricultural drought prediction at downstream of Langat River Basin, Malaysia. Journal of Water and Climate Change, 11(4), 1383–1398.

[16] Homdee, T., Pongput, K., & Kanae, S. (2016). A comparative performance analysis of three standardized climatic drought indices in the Chi River basin, Thailand. Agriculture and Natural Resources, 50(3), 211–219.

[17] Jiang, Z., Rashid, M. M., Johnson, F., & Sharma, A. (2021). A wavelet-based tool to modulate variance in predictors: an application to predicting drought anomalies. Environmental Modelling & Software, 135, 104907.

[18] K. Sundararajan et al., “A contemporary review on drought modeling using machine learning approaches,” CMES-Computer Modeling in Engineering and Sciences, vol. 128, no. 2, pp. 447–487, 2021.

[19] Kanfar, R., Shaikh, O., Yousefzadeh, M., & Mukerji, T. (2020). Real-time well log prediction from drilling data using deep learning. International Petroleum Technology Conference.

[20] Kaur, A., & Sood, S. K. (2020). Deep learning based drought assessment and prediction framework. Ecological Informatics, 57, 101067.

[21] N. Khan, D. A. Sachindra, S. Shahid, K. Ahmed, M. S. Shiru, and N. Nawaz, “Prediction of droughts over Pakistan using machine learning algorithms,” Advances in Water Resources, vol. 139, p. 103562, 2020.

[22] Kim, N., Na, S.-I., Park, C.-W., Huh, M., Oh, J., Ha, K.-J., Cho, J., & Lee, Y.-W. (2020). An artificial intelligence approach to prediction of corn yields under extreme weather conditions using satellite and meteorological data. Applied Sciences, 10(11), 3785.

[23] Kim, T.-W., & Valdés, J. B. (2003). Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. Journal of Hydrologic Engineering, 8(6), 319–328.

[24] Li, G., Hari, S. K. S., Sullivan, M., Tsai, T., Pattabiraman, K., Emer, J., & Keckler, S. W. (2017). Understanding error propagation in deep learning neural network (DNN) accelerators and applications. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 1–12.

[25] Liu, M., & Grana, D. (2019). Accelerating geostatistical seismic inversion using TensorFlow: A heterogeneous distributed deep learning framework. Computers & Geosciences, 124, 37–45.

[26] Lloyd-Hughes, B., & Saunders, M. A. (2002). A drought climatology for Europe. International Journal of Climatology: A Journal of the Royal Meteorological Society, 22(13), 1571–1592.

[27] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5(4), 115–133.

[28] McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. Proceedings of the 8th Conference on Applied Climatology, 17(22), 179–183.

[29] O. Gupta and R. Raskar, “Distributed learning of deep neural network over multiple agents,” Journal of Network and Computer Applications, vol. 116, pp. 1–8, 2018.

[30] O.-J. Jang, H.-T. Moon, and Y.-I. Moon, “Drought Forecasting for Decision Makers Using Water Balance Analysis and Deep Neural Network,” Water (Basel), vol. 14, no. 12, p. 1922, 2022.

[31] Park, S., Im, J., Han, D., & Rhee, J. (2020). Short-term forecasting of satellite-based drought indices using their temporal patterns and numerical model output. Remote Sensing, 12(21), 1–21. https://doi.org/10.3390/rs12213499.

[32] Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., Zhou, L., Liu, H., Ma, Y., & Ding, Y. (2010). The impacts of climate change on water resources and agriculture in China. Nature, 467(7311), 43–51.

[33] Poornima, S., & Pushpalatha, M. (2019). Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. Soft Computing, 23(18), 8399–8412. https://doi.org/10.1007/s00500-019-04120-1.

[34] S. Shamshirband et al., “Predicting standardized streamflow index for hydrological drought using machine learning models,” Engineering Applications of Computational Fluid Mechanics, vol. 14, no. 1, pp. 339–350, 2020.

[35] Qian, Y., Fan, Y., Hu, W., & Soong, F. K. (2014). On the training aspects of deep neural network (DNN) for parametric TTS synthesis. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3829–3833.

[36] Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. Journal of Climate, 23(7), 1696–1718.

An SDN-based Decision Tree Detection (DTD) Model for Detecting DDoS Attacks in Cloud Environment

Jeba Praba. J[1], R. Sridaran[2]
Department of Computer Applications, Christ College, Rajkot, India[1]
Marwadi University, Rajkot, Gujarat, India[1]
Faculty of Computer Applications, Marwadi University, Rajkot, Gujarat, India[2]

***Abstract*—Detecting Distributed Denial of Service (DDoS) attacks has become a significant security issue for various network technologies. This attack has to be detected to increase the system's reliability. Though various traditional studies are present, they suffer from data shift issues and accuracy. Hence, this study intends to detect DDoS attacks by classifying the normal and malicious traffic. The study aims to solve the data shift issues by using the introduced Decision Tree Detection (DTD) model encompassing of Greedy Feature Selection (GFS) algorithm and Decision Tree Algorithm (DTA). It also attempts to enhance the proposed model's detection rate (accuracy) above 90%. Various processes are involved in DDoS attack detection. Initially, the gureKddcup dataset is loaded to perform pre-processing. This process is essential for removing noisy data. After this, feature selection is performed to select only the relevant features, removing the irrelevant data. This is then fed into the train and test split. Following this, Software Defined Networking (SDN) based DTA is used to classify the normal and malicious traffic, then given to the trained model for predicting this attack. Performance analysis is undertaken by comparing the proposed model with existing systems in terms of accuracy, MCC (Matthew's Correlation Coefficient), sensitivity, specificity, error rate, FAR (False Alarm Rate), and AUC (Area under Curve). This analysis is carried out to evaluate the efficacy of the proposed model, which is verified through the results.**

Keywords—Distributed denial of service attack; greedy feature selection; decision tree algorithm; software defined networking; cloud and decision tree detection

I. Introduction

Cloud computing led to the development of technologies due to its services like resource pooling, measured service, broad network access, rapid elasticity, and on-demand self-service. But, security challenges have a dominant issue in cloud computing development. In cloud computing, security requirements include integrity, availability, privacy preservability, confidentiality, and accountability. Among these, availability has been vital, as the main functionality of cloud computing has been to afford on-demand service at various stages. DDoS (Distributed Denial of Service) and DoS (Denial of Service) have been the fundamental techniques to minimize cloud computing availability. SDN (Software Defined Network) has evolved as a novel network paradigm as the SDN characteristics afford the requirement of flexibility, reliability, and secured future networks that could alter the traditional networks. As SDN possesses abilities like decoupling the control plane from the corresponding DP (Data Plane) and centralized controller, traditional networks could be altered with SDN for easy and early detection of DDoS attacks [1]. Clouds and SDNs explore identical designs with three-layered architecture composing an infrastructure layer and computational resources controlled through a control layer that has been controlled through API (Application Program Interface) by applications in the application layer, as presented in Fig. 1.

SDN has been used to deal with DDoS attacks. However, SDN's issues explore the fact that the overall network gets compromised when a specific controller has been flooded with several attacks. The DDoS attack intends to attack the SDN controller through overflowing FT (Flow Table) in the respective DP, as explored in Fig. 2. Cost and limited memory have made FT in DP to be minimum. Hence, whenever a request with an unknown record in FT arises, DP switches forward all the requests directly to the SDN controller, which checks its FT. If it is a legitimate request, it answers with a legal flow. If the requests received have been more at a specific time, the controller takes more time to look at the FT. In addition, the response also enhances that exhausts the controller's resources and makes it unavailable to deal with legitimate requests. Hence, before forwarding a request to the corresponding DP, it is essential to check if the particular request is normal or an attack. In addition, SDN installation would permit novel security issues [2] in SDN as the centralized SDN controller is the bottleneck and turns out to be an SPF (Single Point Failure) in SDN. Attackers will also concentrate on the DDoS attacks, particularly on the SDN controller for flooding the network traffic in the controller. This controller could no longer reply to the elements of DP, like network routers and switches. This collapses the overall network and makes all the cloud services not accessible to end-users through resource exhaustion, resulting in reputation and economic loss.

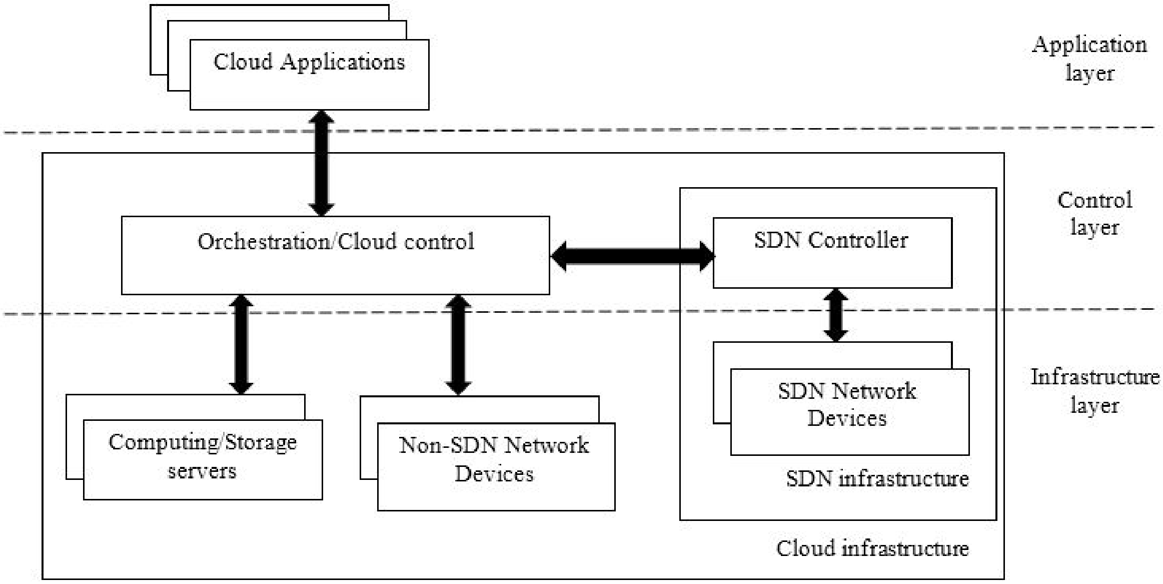


Fig. 1. The Architecture of SDN based Cloud.

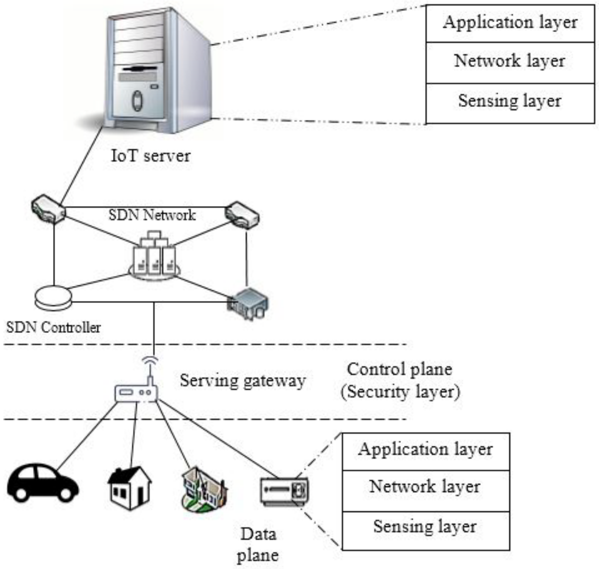


Fig. 2. DDoS Attack Detection in SDN.

Though some efficient studies have been undertaken to detect DDoS attacks in conventional computing environments, these attacks are turning out to be highly occurring in cloud computing environments [3]. The study [4] introduced a design that can detect DDoS attacks in the cloud environment on execution in the real world. Under this design, three ML classification models, RF (Random Forest), LR (Logistic Regression) [5], and SVM (Support Vector Machine), are trained on the CICDDoS2019 dataset. The introduced algorithm chooses the effective classifier for attack prevention, relying on the traffic rate. RF has been found to show better outcomes with 97% accuracy. In addition, a practical and lightweight alleviation method has been introduced for protecting the SDN framework in contradiction to DDoS flooding, confirming an efficient and secure networking environment based on SDN [6]. The proposed system enhanced the DP with a mitigation and classification module to analyze all the new incoming packets and classify all the normal requests from SYN-flood attacks. Subsequently, it performs suitable countermeasures. Simulation outcomes represent that the introduced defending method might effectively deal with DDoS attacks in downstream servers and SDN [7]. The DDoS attack detection in SDN is shown in Fig. 2, which comprises IoT server, SDN controller, serving gateway, data plane, and control plane. Typically, an IoT sensor shall comprise three layers: an application, network, and sensing layer.

The end-user software like email clients and web browsers utilize the application layer. It allows the software to send and receive information and present essential data to the users. The network layer affords the telecommunication resource operations that allow data transfer amongst the systems. PDN (Public Data Network) might provide real-time telecommunication services and have been defined as sub-networks. The sensing layer shows the data type arriving from specific data sources like web services, conventional WSN (Wireless Sensor Networks), and PSNs (Pervasive Social Networks).

Hence, DDoS attack detection is accomplished in SDN, where the SDN controller manages the overall flow control to improve application performance and network management. This platform usually utilizes protocols to inform switches by directing the path for sending the data packets and runs on the server.

Various researches have been carried out to detect DDoS attacks. An approach based on anomaly intrusion has been introduced in the hypervisor layer to minimize the DDoS attacks amongst VMs (Virtual Machines). The evolutionary neural network has been employed to execute the proposed system that integrates PSO (Particle Swarm Optimization) with NN (Neural Network) to detect and classify traffic exchanged amongst VMs. Analysis of the system showed the efficiency in classifying these attacks with high accuracy and a low false alarm rate [8]. However, the dataset used handles only the traffic exchange amongst VMs; hence, traffic arriving from the outside host could be researched in further studies. As this system relied on soft computing methods, a probable future work has to be chosen as the alternative technique to accomplish a high detection rate and minimum computation time. Similarly, ML (Machine Learning) based system has been recommended to detect DDoS attacks. This system makes inferences by the signatures extracted earlier from network traffic samples. Experimentations have been performed through four benchmark datasets, and the outcomes exhibited an accuracy rate of 96.5%. However, the accuracy rate has to be further improved to enhance the detection rate [9], [10]. However, accuracy and dataset shift issues are common in existing methods. Hence, the present study intends to detect DDoS attacks in the SDN-based cloud environment through the proposed Decision Tree Detection (DTD) model.

The objectives of this study are 1) to address the dataset shift issue efficiently through the proposed Decision Tree Detection (DTD) model comprising Greedy Feature Selection (GFS) algorithm and Decision Tree Algorithm (DTA). 2) To enhance the detection accuracy above 90% using the introduced Decision Tree Detection (DTD) model. 3) To evaluate the performance of the proposed system by comparing it with traditional ML methods concerning detection rate, error rate, FAR (False Alarm Rate), specificity, sensitivity, MCC (Matthew's Correlation Coefficient), and AUC (Area under Curve).

A. The Significant Contribution of the Proposed Study

The proposed model contributes to increasing the security in the cloud system with high accuracy. The proposed system employed GFS in the feature selection process, which reduces the overfitting issues, the accuracy range was improvised, and the model training was accomplished faster. The proposed system then used a DTD algorithm for classification, which forms in a tree structure. Further, it breaks down the dataset into smaller subsets, whereas the related decision tree is incrementally developed at the equivalent period. The system also efficiently performs data shifting. Various studies implemented with DTA and GFS in cloud environments were viewed. Therefore, the system can effectively and efficiently detect DDoS attacks in the SDN-based cloud system.

B. The Motivation of the Proposed Model

Though various studies have been implemented in the DDoS attack detection over the SDN-based cloud system with the DT approach, they still fail to enhance the accuracy range. The Decision tree provides an efficient approach for decision-making because it results in lay for the problems, thus that all choices might be challenged. Consents to estimate all the conceivable magnitudes of a decision. Affords an architecture for enumerating the standards of results and the possibilities of accomplishing them. Then, the Greedy algorithm results elucidation for small illustration of the complications can be forthright and easy to understand. The greedy feature selection process results in both forward and backward selection; in the proposed model, the backward selection is employed. Thus, these factors motivate implementation of an SDN Based Decision Tree Detection (DTD) Model for Detecting DDoS Attacks in a Cloud Environment.

C. Paper Organization

The paper is organized in the following way. Section I explores the SDN-based cloud's basic ideas in DDoS attack detection. Followed by the existing works related to this context are analyzed in Section II. Then, the proposed DTD model is described in Section III. The results obtained from the proposed system are discussed in Section IV. Finally, the overall study is summarized in Section V.

II. Review of Existing Work

Various existing studies that correspond to DDoS (Distributed Denial of Service) attack detection in SDN-based cloud are analyzed, and the outcomes are presented. The common problems encountered in these existing systems during this review are also explored in this section.

The SDN (Software Defined Networking) encompasses various management, control, and configuration functionalities from the server. This SDN is partitioned into the control and DP method. From the recent data centers that maintain massive data every moment with servers possessing large data volumes, the SDN driver has increased. Components needed in this technique are costly. This process also consumes more time, and configuration turns out to be manual. Through central and management control, this issue has been solved in SDNs. However, SDN has a weak structure in various threats where DDoS attacks dominate. The main recent disruptions in all security systems have been because of DDoS attacks. These attacks mainly intend to collapse the user's access path to another server or network source. It occurs by integrating the server and host. Because of this, resources like CPU, memory, and traffic disappear from the host. Hence, this issue occurs when data transfer happens between authentic users and servers. Supplementary attacks could be resolved through rebooting. However, this merging model or flooding is complex. The DDoS attack detection has been analyzed based on entropy in SDN for detecting and controlling the impact of the SDN controller. DDoS attack traffic has been incorporated into typical traffic by a setup of twenty-five and fifty percent of traffic intended towards a host in the SDN network. Simulation has been carried out, and the outcomes explore the threshold value selected to find an effective DDoS attack. Future studies include simulating chi-square to find the attack traffic incorporated with the typical traffic [11]. Similarly, a method based on $\mathrm{Inf}_{\mathrm{dis}}$ (Information distance) has been used to detect this attack in SDN based cloud environment. Subsequently, ABA (Adaptive Boosting Algorithm) framework has been employed with SDN features to detect DDoS attacks. Finally, experimentations revealed the effectiveness of ABA in detecting this attack in SDN based cloud. Despite various merits in utilizing SDN in the cloud, it makes the cloud system susceptible to multiple novel security attacks like FTO (Flow-Table Overloading) DDoS attacks [12]. The FTS (Flow Table Sharing) method has been used to protect SDN-based clouds from FTO DDoS attacks to prevent this attack. This method uses idle FT of supplementary OFS (Open Flow Switches) in-network to protect the FT of switches from overloading. The proposed method enhances the cloud system's resistance against DDoS attacks with minimum engagement of the SDN controller. This leads to minimum communication overhead. The proposed approach has been highly supported by many experiments based on simulation. This shows the efficacy of the proposed system. It is also significant to classify abnormal and normal traffic [13].

Ensemble classification methods comprising SVM (Support Vector Machine), ELM (Extreme Learning Machine), and K-Nearest Neighbour (K-NN) have been suggested for the detection of DDoS attacks through the classification of the traffic as abnormal and normal. The analytical results explored that the proposed K-NN shows an accuracy rate of 76.9%. ELM and SVM classifier performs less or more identical to one another, with an accuracy of 96.4% and 92.7%. The overall decision is undertaken through a max-vote methodology [14], [15]. Various existing methods used different techniques for detecting this attack. The distributed blockchain method has been employed to detect and prevent DDoS attacks on SDN's centralized control plane. The proposed system has been simulated through the use of the AnyLogic simulator. The outcomes revealed the efficiency of the introduced system more than traditional systems, as it adds only minor overhead. Results explored that the controller's overhead was minimized up to thirty-five percent. This also substantially minimized the SDN controller's DDoS attack risk and overhead. A HIDS (Host-based Intrusion Detection System) has been presented to monitor the intrude's activity. The host machine would permit the administrator to monitor the attacker and their activities and alert the data owner in the cloud [16]. The proposed method has enhanced efficacy over the overall system's performance [17], [18].

Similarly, TEHO-DBN (Taylor Elephant Herd Optimization-Deep Belief Network) has been used to detect DDoS attacks in the cloud computing environment. This proposed classifier determines if the particular user is normal or an attacker. Simulation has been undertaken, and it could be summarized that the introduced TEHO relying on DBN has enhanced the performance with an accuracy of 83%. Though the accuracy is better, it has to be further improvised for efficient detection of DDoS attacks [19]. Hence, a Bi-fold SDN-based solution has been recommended using a covariance matrix and genetic algorithm (GA). Traffic data (real-time) has been gathered from an analyzer tool named Tshark network. The Bi-fold method has been employed to distinguish the abnormal traffic. GA takes an initial decision regarding the abnormal and normal attacks. The covariance matrix has been used for refining decisions. Empirical outcomes confirmed the efficiency of the introduced method with better sensitivity, specificity, and accuracy. But, the consumption of time to detect attacks is higher, but it is tolerable simultaneously. In addition, minimizing biased data is also significant in enhancing the attack detection rate [20].

The article [21] examined all the features extracted from SDN traffic, minimizing bias data from the dataset. The traffic features have been assessed through a tenfold-cross validation method. The efficiency of the proposed dataset has been assessed through comparison with the supplementary dataset, for instance-KDDCUP99 (Knowledge Discovery and Data mining tools Competition) dataset. The outcomes revealed that the introduced dataset could be efficiently used for SVM on SDN. A live traffic analysis technique has been provided with the NN (Neural Network) [22]. The proposed TFC-NN (Traffic Flow Classifier-Neural Network) has been trained by a labeled dataset built from under traffic and regular traffic of SDN. A live reduction process has also been integrated with TFC-NN relying on detecting DDoS. The recommended method has been deployed and assessed on SDN architecture relying on various performance metrics under different scenarios of DDoS attacks. Through TFC-NN, Classification has been accomplished with Global accuracy (96.13%). SDN and fog computing has been integrated as a mitigation method to accomplish better outcomes [23]. It also considers the IP spoof an excellent way to detect DDoS attacks. Proposed IP-spoof detection has been undertaken near the attacker source in this study to enhance the attack trace. In addition, a model has been introduced for detecting and mitigating all the DDoS attacks in the cloud environment [24], [25]. The introduced model needs small storage and the ability for fast detection. Empirical outcomes explored the power of the system to ease many attacks. Processing time and detection accuracy were the performance metrics used to assess the proposed model's performance. From the outcomes, it has been clear that the proposed system accomplished high accuracy of 97% with reduced false alarms [26], [27]. Similarly, issues have been solved by introducing an effective system named Prodefense to detect and mitigate DDoS attacks. It also includes criteria that are application-specific for the corresponding threshold of the network traffic. This allows the execution of customizable measures to detect DDoS attacks [28].

Likewise, a modular and flexible architecture has been suggested to alleviate and detect LR-DDoS (Low Rate DDoS) attacks in SDN-based clouds. Notably, the IDS (Intrusion Detection System) has been trained in the suggested architecture through the use of six ML (Machine Learning) models such as RF (Random Forest), SVM, MLP (Multilayer Perceptron), RT (Random Tree), and J48 to assess their performance through the use of CIC (Canadian Institute of Cybersecurity) DoS dataset. Evaluation findings reveal that the introduced method accomplished a 95% detection rate, irrespective of the complexity of detecting LR-DoS attacks. Simulation has been carried out equivalent to real-world production through the usage of the ONOS (Open Network Operating System) controller that has been running on MVM (Mininet Virtual Machine), which showed better outcome. Fast attack detection is also another significant parameter to be considered [29]. For this purpose, a DDoS attack alleviation architecture has been recommended to combine a programmable network observance to permit flexible controlling structure and attack detection for specific and fast attack reactions. To manage the structure, an attack detection system based on a graphic model has been introduced that could handle the dataset shift issue [30]. Simulation outcomes revealed that the suggested architecture could efficiently and effectively solve the security issues caused by the novel network paradigm. It has also been concluded that the proposed attack detection could efficiently state several attacks through real-world cases. Empirical analysis of ML methods has been carried out for detecting Botnet DDoS attacks [31].

Evaluation has been carried out on KDD99 and UNBS-NB 15 datasets for detecting Botnet DDoS. Typically, ML methods such as SVM, NB (Naïve Bayes), USML (Unsupervised Machine Learning), ANN (Artificial Neural Network), and DT (Decision Tree) have been analyzed concerning FPR (False Positive Rate), AUC (Area under Curve), accuracy, FAR (False Alarm Rate) and MCC (Matthews Correlation Coefficient). Analytical results revealed that the KDD99 dataset's performance was better than UNBS-NB 15. This substantiation has been crucial in network security and other relevant areas [32].

Various problems identified through the review of different existing methods for DDoS attack detection in SDN-based cloud environments are discussed below,

- Only a few parameters have been taken into analysis that encompasses the FPR (False Positive Rate) and attack detection rate [12], [16]. In addition, the current work [14], [20] considered only accuracy, specificity, and sensitivity. The present study considers many performance metrics for comparative analysis, such as detection rate, error rate, FAR (False Alarm Rate), specificity, sensitivity, MCC (Matthew's Correlation Coefficient), and AUC (Area under Curve), which explores the effective analysis of the proposed system.

- The traditional research [14] used K-NN, ELM, and SVM for DDoS attack detection, and the accuracy rate of K-NN was found to be 76.9%, ELM-96.4%, and SVM-92.7%. The existing system [19] used TEHO-DBN to detect this attack, and the accuracy was 83%. In addition, the article [22] accomplished an accuracy rate of 96.13% through the use of the proposed TFC-NN. Similarly, the paper introduced methods like HIDS with an accuracy of 97 per cent. Only two parameters are considered [24]. Likewise, the article [29] used ML-based methods, and accuracy was 95 per cent. However, accuracy has to be improved further in all these cases for efficient DDoS attack detection. Hence, this article intends to improve the detection accuracy through the proposed DTD model. Its efficiency is confirmed through the results.

- The existing work [22] has not executed and deployed the NIDS (Network-Based Intrusion Detection System) in SDN. However, the present study intends to detect DDoS attacks in SDN based cloud environment.

- Introduced model of the traditional systems performs fast execution. But, these works hardly suffer from performance loss regarding dataset shift issues [30] and detection accuracy [31]. The proposed system aims to solve these issues through the proposed DTD model.

- The traditional system [32] aimed to compare the introduced methods with other ML methods through many evaluation metrics in the future. But, the present work performs comparative analysis in terms of several metrics by considering various ML methods, such as SVM, NB, DT, USML and deep learning (DL) algorithm - ANN.

- Though various studies [31], [32], [28], [23] have been implemented in the DDoS attack detection, it fails to focus on the data shifting issues.

III. Proposed Methodology

The research introduced a model named DTD (Decision Tree Detection), comprising two algorithms such as Greedy Feature Selection (GFS) and Decision Tree Algorithm (DTA), to detect DDoS attacks in SDN based cloud environment. Various techniques and methods exist to detect this attack. In the proposed model GFS is employed for feature selection, which reduces the complexity and selects features faster. Followed by this, the classification is performed with DTA that enhance the accuracy range in class. However, all these methods possess common drawbacks. The data shift issues are not efficiently handled, and detection accuracy is also minimal [30]. To resolve this drawback, this study proposed a DTD model comprising two algorithms such as GFS and DTA, where GFS is used to perform feature selection. For this purpose, the gureKddcup dataset is used, which includes 48 features. In addition, DTA (Decision Tree Algorithm) is used for classification. The overall process of the proposed system is presented in Fig. 3.

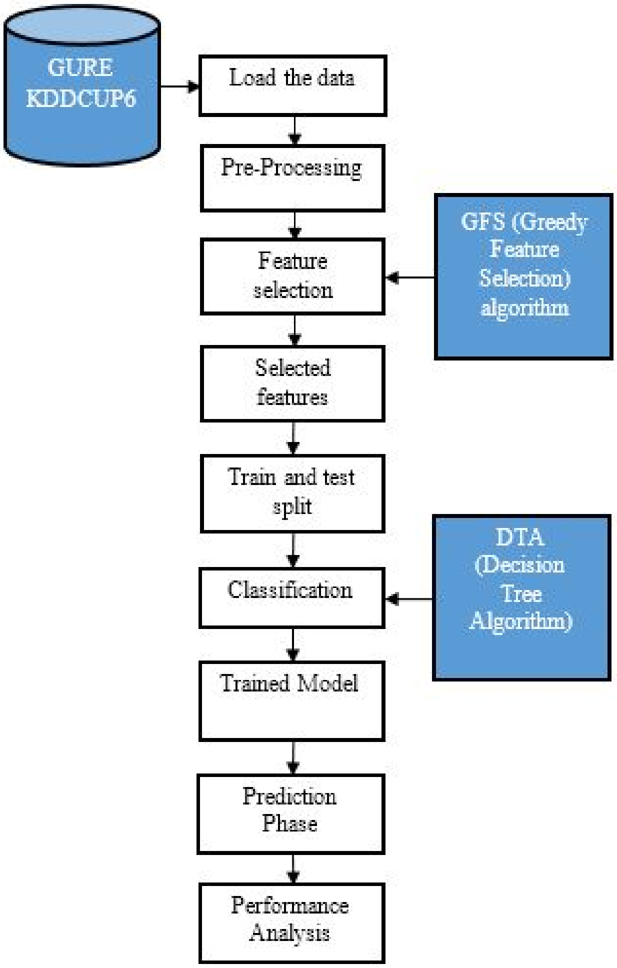


Fig. 3. Dataflow Diagram of the Proposed DTD Model.

Various processes are involved in detecting DDoS attacks. At first, the dataset is loaded, and then pre-processing is performed for noise removal. After this, the feature selection is performed by the GFS algorithm, which filters only five features from all the features available in the dataset; these features are then fed into DTA, one of the efficient classification algorithms that classify non-malicious (normal) and malicious attacks. Here, one portion of the dataset is used for training. The other portion is used for testing. Training data is labeled as local. At the same time, the testing data is labeled as global. After training, the proposed model can detect DDoS attacks effectively, which is proven through results. Finally, performance analysis is undertaken to evaluate the efficiency of the proposed system.

A. GFS (Greedy Feature Selection) Algorithm

GFS is a mathematical method that is simple, easy to implement, and provides solutions to complex issues by making practical decisions [33]. This algorithm operates by recursively building object sets from minimum probable constituent elements. It either selects the best features individually (forward selection) or removes all the worst features individually (backward selection). In the proposed model, the backward selection is employed. The dataset is loaded for pre-processing and feature selection using the GFS algorithm. It performs various steps to accomplish this process, and those steps are presented,

Step 1: Initialize the dataset and its source with the attack features.

Step 2: Generate the objects for the evaluator, search algorithms, and attribute selection.

Step 3: Initiate a Greedy backward search with a filter in accordance with the search algorithms and evaluator over a particular dataset.

Step 4: Measure the error of LOOCV (Leave One Out Cross Validation) of DT classification for the current population set of features of the current search iteration. This is the fitness cost $f(x_{iter})$ for the input set of features of the current iteration x_{iter}.

Step 5: Apply filter to perform greedy operations in a stepwise way. Evaluate it to optimize the search filters.

Step 6: Obtain the count of classes and their attributes. Map the class indexes and update their weight for a pre-defined count of instances.

Step 7: Repeat step 2 to 7 for maximum search iterations.

Step 8: Save and update the minimum fitness cost as $f_{min}(x) = \min f(x)$.

Step 9: Final optimal solution of GFS comprises the significant features from the dataset.

The fitness cost $f(x)$ serves as the parameter to decide on the selection of significant features since it is the evaluation measure of the feature set x. The GFS is executed until the $f_{min}(x)$ is obtained as the optimal solution. Hence, after implementing all these steps through GFS, the dataset comprising 48 columns gets reduced to 5 columns with only relevant and specific features. Feature selection must be made efficiently as it affects the accuracy rate. Implementing GFS for feature selection undergoes the seven steps to select the best features effectively. Finding an issue's solution is typically easier with the GFS algorithm than with other algorithms. Hence, implementing it will enhance the detection accuracy proven through outcomes.

B. DTA (Decision Tree Algorithm)

DTA pertains to the group of SLA (Supervised Learning Algorithms) [34]. It is a tree-based classifier where the internal nodes show the dataset features, individual leaf nodes show the results, and branches show decision rules. Unlike other SLAs, DTA can be used for classification and regression issues. DTs are efficient kinds of algorithms that rely on several learning techniques. It possesses various advantages by boosting the accuracy of prediction models, stability, and straightforward interpretation. DTA is an efficient classification algorithm that exhibits data records into corresponding classes. It utilizes the recursive partition method for data exploration. The DTA components include roots, leaves, and branches. This study pursues a directed tree, which means roots don't have edges. Other components possess a single edge. In addition, the interior nodes show nodes without flow edges. Other nodes are leaves which show the decisional or terminal nodes. The Interior node partitions the space decision into many subspaces based on minimized feature sets. As numeric attributes are considered, attribute spaces are termed conditional ranges. Leaf nodes hold target values to attain their corresponding classes. Under the conditional values, the arrangement of interior nodes occurs from the root node to the leaf nodes. Hence, DTA is used for feature classification based on the below steps.

Step 1: Initialize the tree with a root node (R) that comprises the overall dataset.

Step 2: Determine all the best attributes in the dataset through ASM (Attribute Selection Measure).

Step 3: Determine all the best attributes in the dataset through ASM (Attribute Selection Measure) using information gain I. Information gain is the criterion for estimating the information comprised by each feature attribute. It can be expressed as,

$$I = E_R - (A_R * E_x)$$

Where, E_R is the entropy of the dataset, E_x is the entropy of the feature x, and A_R is the weighted average of the dataset. This measure of entropy helps in the identification of redundant or unnecessary information in an attribute and in the specification of randomness in the data.

Step 4: Generate the node of DT that comprises the best attributes.

Step 5: Recursively make new DTs using dataset subsets developed in step 3. Iterate this process until a particular stage is met, where further classification of nodes cannot be performed. These final nodes are the leaf nodes.

Hence, all the features selected using the GFS algorithm are fed into DTA, which classifies the features based on the above five steps. The attacks are classified as malicious or non-malicious through the overall proposed DTD model. The efficacy of the proposed system is confirmed through the outcomes, which are discussed in the subsequent section.

IV. Results and Discussion

A. Experimental Setup and Dataset Description

The research considered the dataset to evaluate the proposed model for detecting DDoS attacks. It is described in this section. The proposed system also attempts to solve the data shift issue, a common problem in the existing system. A comprehensive analysis of data shift is also explored in this section.

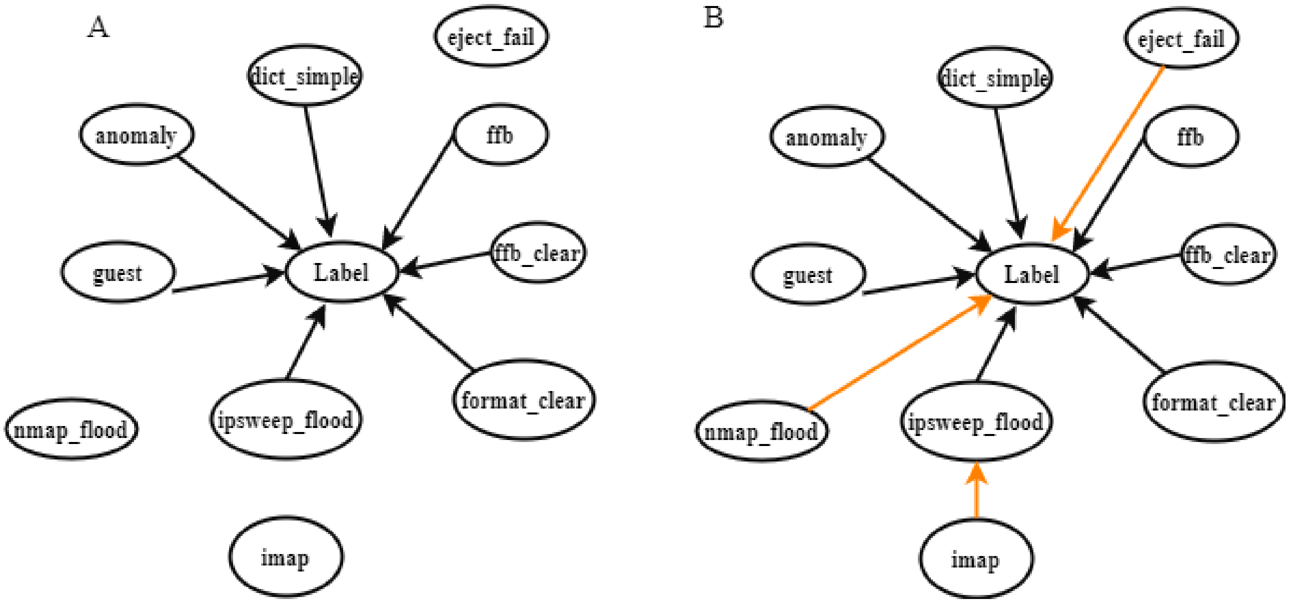
1) Experimental setup: The proposed system is developed and implemented in a system having configurations, as shown below.

Hardware details of the introduced module: Windows 10 pro processor: Intel (R) Core ™ i5-4210U CPU @1.70GHz installed memory, RAM: 4GB, System type: the 64-bit operating system.

2) Dataset description: This study used the gureKddcup dataset [35] to assess the performance of the proposed algorithms, which is one of the widely utilized datasets. It is created regarding the association of Kddcup99 (UCI repository database) and incorporates its payload into individual connections. This dataset helps extract all the information directly from a separate connection's payload to be effectively used in ML processes. In this study, the gureKddcup dataset is used to detect DDoS attacks that consist of 48 attributes which are later reduced to five through the proposed GFS algorithm.

3) Data shift: The data shift issue handles the information association in two subsets of data and assists in predicting a subset, thereby taking into account the data in the supplementary subset. This issue happens when the data generation relies on a model P y1/x1 P(x1), where P(x1) indicates the data distribution or changes amongst the train and test split. It usually occurs when data from a particular class is selected spontaneously compared to a supplementary class. Hence, a large dataset is needed to accomplish high accuracy in this case. During data classification and dataset partitioning in training and testing split, the training dataset is termed Local. The testing dataset is termed Global. It is shown in Fig. 4. As per Fig. 4, it could be seen that a data shift issue occurs in network traffic when a model is constructed using a training dataset. In the proposed method, when new traffic arrives, only some existing data is used as training data (Global). This shows that the proposed detection model keeps updating its training data according to the data received in real-time. Hence, it could always get new observations, afford accurate outcomes, and solve data shift problems. The proposed model nearly functions in real-time and thus solves data shift issues. This infers that the proposed model is not limited or constrained to any particular dataset. To prove the robustness of the model, its performance is compared with the performance of existing datasets.

4) Performance metrics: The performance of the proposed system is analyzed concerning detection rate (accuracy), error rate, specificity, sensitivity, FAR (False Alarm Rate), AUC (Area under Curve), and Matthew's Correlation Coefficient (MCC). Each of the performance metrics is discussed in this section.



(A): Local Update

(B): Global Update.

Fig. 4. Comprising A and B, explores how the Introduced Model Efficiently Handles Data Shift Issues. It is also essential as it affects the Proposed Model's Accuracy. Hence, the Proposed System Efficiently Solves Data Shift Issues, enhancing the Accuracy Rate above 90%.

Accuracy in detecting the attacks can be described as the proportion of detected attacks to the overall attack counts. It is given by the following equation 1.

$$\text{Accuracy} = \frac{\text{Count of detected attacks}}{\text{Overall attack count}} \qquad (1)$$

a) Error rate: The error rate is the proportion of attack counts not detected by the overall attack count and is given by equation 2.

$$\text{Error rate} = \frac{\text{Count of attacks not detected}}{\text{Overall attack count}} \qquad (2)$$

b) Sensitivity: It is defined as the number of true positives correctly predicted and given by equation 3.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive+False Positive}} \qquad (3)$$

c) Specificity: The state/quality of being specific/unique to an individual or a group. It is mathematically represented as per equation 4.

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative+False Negative}} \qquad (4)$$

d) AUC (Area Under Curve): It could be stated as the correct curve integral that explores differences in classification and is given by equation 5.

$$\text{AUC} = \frac{1}{2}((\frac{\text{True Positive}}{\text{True Positive+False Negative}})+(\frac{\text{True Negative}}{\text{True Negative+False Positive}})) \qquad (5)$$

e) FAR (False Alarm Rate): It is defined as the ratio in which FA occurs in contradiction to true alarms.

f) MCC (Matthew's Correlation Coefficient): A highly static and reliable rate would afford a high score if detection attained good outcomes in all four confusion matrix classes (True Positive, False Positive, True Negative, and False Negative). It is given by equation 6.

$$\text{MCC} = \frac{\text{True Positive}*\text{True Negative}-\text{False Positive}*\text{False Negative}}{\sqrt{\text{(True Positive+False Positive)(True Positive+False Negative)(True Negative+False Positive)(True Negative+False Negative)}}} \qquad (6)$$

B. Experimental Results

The proposed DTD model is implemented, and the results are shown in this section. Initially, the dataset is uploaded. The dataset used in the study is gureKddcup. After the dataset is uploaded, the dataset is viewed. The overall count of instances or records is 10000, and attributes (features) are found to be 48. The proposed evolution of the proposed model is stated in Table I with respect to the class and considered performance metrics. Initially, in class normal, the true positive is 0.985, the false positive is 0.043, Precision is 0.999, recall is 0.985, F-measure is 0.992, and the ROC area is 0.972. In the class warezclient, true positive is 0.965; true negative is 0.015, Precision is 0.428, recall is 0.965, F1-measure is 0.593, and ROC area is 0.996. Similarly, in the class dict, the true positive is 0.961, the true negative is 0, Precision is 0.98, recall is 0.961, F1-measure is 0.97, and ROC area is 0.999. Therefore, in class warezmaster, the true positive is 0, the true negative is 0, Precision is 0, recall is 0, F1-measure is 0, and ROC area is 1. For class teardrop, the true positive is 0.982, the true negative is 0, Precision is 1, recall is 0.982, F1-measure is 0.991, and ROC area is 1. In syslog class, the true positive is 0, the true negative is 0, Precision is 0, recall is 0, F1-measure is 0, and ROC area is 0.499. Similarly, in land class, the true positive is 0, the true negative is 0, Precision is 0, recall is 0, F1-measure is 0, and ROC area is 0.499. In the guest class, the true positive is 0, the true negative is 0, Precision is 0, recall is 0, F1-measure is 0, and the ROC area is 0.489. Similarly, in class imap, the true positive is 0, the true negative is 0, Precision is 0, recall is 0, F1-measure is 0, and ROC area is 0.489. In the weighted average class, the true positive is 0.984, the true negative is 0.043, Precision is 0.992, recall is 0.984, F1-measure is 0.987, and ROC area is 0.976.

The confusion matrix is the performance evaluation of ML classification issues. The output could be two or many classes, a table with four varied combinations of actual and predicted values. The confusion matrix for the proposed model is done as shown in Table II. As per Table II, the diagonal values show the correct prediction rate. Finally, the correct and incorrect classified instances are determined as per Table III.

From Table III, the correctly classified instances are 98.42%, and the incorrectly classified instances are 1.58%. In addition, the kappa statistics, MAE (Mean Absolute Error), RMSE (Root Mean Square Error), RAE (Relative Absolute Error), and RRSE (Root Relative Squared Error). The experimental results show that the proposed system shows high accuracy with a low error rate. The proposed system is compared with the existing system to prove the efficacy of the introduced system over other systems, which is explored in the next section.

TABLE I. PERFORMANCE ANALYSIS

TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
0.985	0.043	0.999	0.985	0.992	0.976	normal
0.965	0.015	0.428	0.965	0.593	0.996	warezclient
0.961	0	0.98	0.961	0.97	0.999	dict
0	0	0	0	0	1	warezmaster
0.982	0	1	0.982	0.991	1	teardrop
0	0	0	0	0	0.499	syslog
0	0	0	0	0	0.499	land
0	0	0	0	0	0.489	guest
0	0	0	0	0	0.489	imap
0.984	0.043	0.992	0.984	0.987	0.976	Weighted Average

TABLE II. CONFUSION MATRIX

a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**<--classified as**
9627	142	1	0	0	0	0	0	0	a=normal
4	110	0	0	0	0	0	0	0	b=warezclient
2	0	49	0	0	0	0	0	0	c=dict
0	3	0	0	0	0	0	0	0	d=warezmaster
1	0	0	0	56	0	0	0	0	e=teardrop
0	1	0	0	0	0	0	0	0	f=syslog
1	0	0	0	0	0	0	0	0	g=land
2	0	0	0	0	0	0	0	0	h=guest
0	1	0	0	0	0	0	0	0	i=imap

TABLE III. CORRECTLY AND INCORRECTLY CLASSIFIED INSTANCES

Correctly classified instances	98.42
Incorrectly classified instances	1.58
Kappa statistics	0.7281
Mean absolute error	0.0035
Root mean squared error	0.0587
Relative absolute error	34.0313
Root relative squared error	82.7176
Total number of instances	10000

C. Comparative Analysis

The proposed system is analyzed by comparing it with other algorithms concerning seven performance metrics. The existing algorithms considered for analysis include SVM (Support Vector Machine), DT (Decision Tree), NB (Naïve Bayes), ANN (Artificial Neural Network), and USML (Unsupervised Machine Learning). At first, the comparison is made by considering the existing and proposed model's detection rate and error rate over the corresponding input dataset. It is shown in Table IV.

TABLE IV. COMPARATIVE ANALYSIS OF THE PROPOSED AND TRADITIONAL SYSTEMS [30]

Parameter	**Local**		**Global**	
	[30]	**DTD (Proposed work result)**	**[30]**	**DTD (Proposed work result)**
Detection rate	86.56%	88.80%	89.30%	98.42%
Error	13.44%	11.20%	10.70%	1.58%

From the comparative analysis, as shown in Table IV, it is found that the proposed DTD model shows an 88.80% detection rate in local data and an error rate of 11.20%. In comparison, the existing system [30] shows a detection rate of 86.56% and an error rate of 13.44% in local data. The proposed DTD shows a detection rate of 98.42% and an error rate of 1.58% in Global data. In contrast, the traditional system [30] offers an 89.30% detection rate and 10.70% error rate in Global data. It is also graphically presented in Fig. 5. Hence, the introduced model affords more effective accuracy than traditional systems in both local and global data. In real-time, the attackers don't send similar attack patterns each time, which might vary. As the experimental outcomes show the efficiency of the proposed system in testing data with high accuracy, the introduced model can also be utilized in real-time.

In addition, a comparative analysis is undertaken by comparing the proposed DTD model with the existing methods, such as SVM, NB, ANN, DT and USML in terms of accuracy, sensitivity, specificity, FAR, AUC and MCC. The obtained results are shown in Table V and Fig. 6.

The comparative analysis explores that the proposed DTD model shows an accuracy rate of 98.42%, existing SVM shows a 91.55% accuracy rate, DT of 93.30%, NB of 96.74%, ANN of 97.44%, and USML offers 98.08%. From this comparison, the proposed system shows high accuracy rate than the existing systems. The proposed system also shows high AUC, MCC, sensitivity, and specificity than the traditional systems. The FAR of the proposed system is 1.58%, which is minimum than the existing methods, revealing that the introduced system shows only a minimum error rate with high accuracy. Hence, it can be concluded that the presented system is more effective than the traditional systems in terms of all the considered metrics.

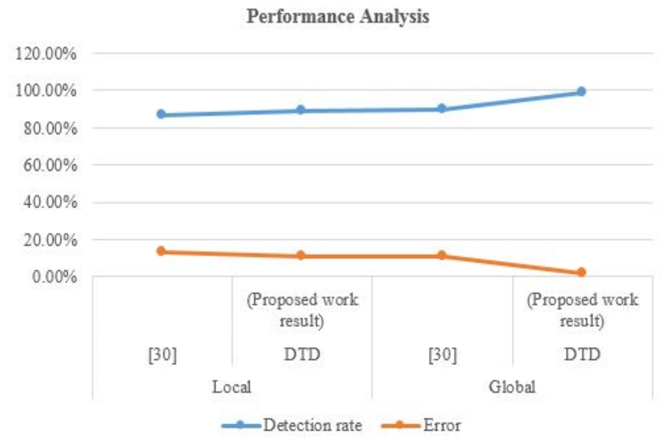


Fig. 5. Comparative Analysis in Terms of Detection Rate and Error Rate [30].

TABLE V. COMPARATIVE ANALYSIS IN TERMS OF VARIOUS METRICS [31]

Performance metrics	**SVM**	**DT**	**NB**	**ANN**	**USML**	**DTD (Proposed work result)**
Accuracy	91.55%	93.30%	96.74%	97.44%	98.08%	98.42%
FAR	8.45 %	6.70 %	3.26 %	2.56 %	1.92 %	1.58%
Sensitivity	90.13%	93.14%	98.21%	84.89%	91.88%	98.84%
Specificity	9.87 %	6.86 %	1.71 %	15.11%	8.12 %	94.30%
MCC	10.46%	5.48 %	10.42%	14.46%	1.48 %	90.26%
AUC	89.54%	94.52%	89.58%	85.54%	98.52%	98.90%

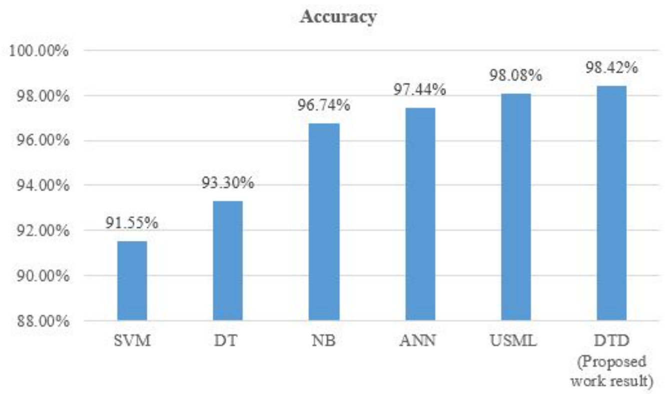


Fig. 6. Comparative Analysis in Terms of Accuracy [31].

V. Conclusion

The study detected DDoS (Distributed Denial of Service) attacks through the use of the proposed DTD (Decision Tree Detection) model that is composed of the GFS (Greedy Feature Selection) algorithm for selecting relevant features and DTA (Decision Tree Algorithm) for classifying these features. The proposed model was assessed by comparing it with traditional algorithms such as SVM (Support Vector Machine), DT (Decision Tree), NB (Naïve Bayes), ANN (Artificial Neural Network), and USML (Unsupervised Machine Learning) concerning significant metrics such as accuracy, MCC, sensitivity, specificity, error rate, FAR and AUC. The outcomes explored that the proposed system showed a high accuracy of 98.42% in testing data. As the proposed system showed high accuracy in testing data, it can be employed in real-time and is expected to get efficient results in detecting DDoS attacks. The proposed approach is also more effective than traditional methods, which are confirmed through the outcomes. Hence, these merits show the efficacy of the proposed system in classifying the malicious and normal attacks, thereby efficiently predicting. This results in the proposed model being employed in real-time to enhance the security in the cloud environment. The proposed model is evaluated with only Gurekddcup6 dataset; it can also be evaluated with other datasets. In the future, various other algorithms and hybrid approaches can improve detection accuracy further and perform DDoS attack mitigation.

References

[1] F. Shaar and A. Efe, “DDoS attacks and impacts on various cloud computing components,” International Journal of Information Security Science, vol. 7, no. 1, 2018.

[2] K. Bhushan and B. B. Gupta, “Security challenges in cloud computing: state-of-art,” International Journal of Big Data Intelligence, vol. 4, no. 2, pp. 81-107, 2017.

[3] Y. Cui et al., “Towards DDoS detection mechanisms in Software-Defined Networking,” Journal of Network and Computer Applications, p. 103156, 2021.

[4] P. O. Prakash, K. Sasirekha, and D. Vistro, “A DDOS prevention system designed using machine learning for cloud computing environment,” International Journal of Management (IJM), vol. 11, no. 10, 2020.

[5] Y. BN, “Preemptive modelling towards classifying vulnerability of DDoS attack in SDN environment,” International Journal of Electrical & Computer Engineering (2088-8708), vol. 10, no. 2, 2020.

[6] S. Mahrach and A. Haqiq, “DDoS Flooding Attack Mitigation in Software Defined Networks,” International Journal of Advanced Computer Science and Applications, vol. 11, no. 1, pp. 693-700, 2020.

[7] T. Ubale and A. K. Jain, “Survey on DDoS attack techniques and solutions in software-defined network,” in Handbook of computer networks and cyber security: Springer, 2020, pp. 389-419.

[8] A. Rawashdeh, M. Alkasassbeh, and M. Al-Hawawreh, “An anomaly-based approach for DDoS attack detection in cloud environment,” International Journal of Computer Applications in Technology, vol. 57, no. 4, pp. 312-324, 2018.

[9] F. S. d. Lima Filho, F. A. Silveira, A. de Medeiros Brito Junior, G. Vargas-Solar, and L. F. Silveira, “Smart detection: an online approach for DoS/DDoS attack detection using machine learning,” Security and Communication Networks, vol. 2019, 2019.

[10] R. Kesavamoorthy and K. R. Soundar, “Swarm intelligence based autonomous DDoS attack detection and defense using multi agent system,” Cluster Computing, vol. 22, no. 4, pp. 9469-9476, 2019.

[11] R. Chaganti, “Review of Distributed Denial of Service Attack Detection Techniques in Software Defined Networking and Cloud Computing.”

[12] S. W. Tufa, M. Mengstie, H. Gebregziabher, and B. R. Babu, “Detecting Ddos Attack Using Adaptive Boosting with Software Defined Network in Cloud Computing Environment,” REVISTA GEINTEC-GESTAO INOVACAO E TECNOLOGIAS, vol. 11, no. 4, pp. 3485-3494, 2021.

[13] K. Bhushan and B. B. Gupta, “Distributed denial of service (DDoS) attack mitigation in software defined network (SDN)-based cloud computing environment,” Journal of Ambient Intelligence and Humanized Computing, vol. 10, no. 5, pp. 1985-1997, 2019.

[14] S. Vimala and J. Dhas, “SDN based DDoS attack detection system by exploiting ensemble classification for cloud computing,” International Journal of Intelligent Engineering and Systems, vol. 11, pp. 282-291, 2018.

[15] A. S. Alzahrani, “An Optimized Approach-Based Machine Learning to Mitigate DDoS Attack in Cloud Computing,” International Journal of Engineering Research and Technology, vol. 13, no. 6, pp. 1441-1447, 2020.

[16] J. Vinnarasi and N. Sudha, “Security Solution for SDN Using Host-Based IDSs Over DDoS Attack,” International Journal of Emerging Technology and Innovative Engineering, vol. 5, no. 9, 2019.

[17] T. Pandikumar and T. Belissa, “Distributed Denial of Service (DDOS) Attack Detection in Software Defined Networking with Cloud Computing,” International Journal of Engineering Science, vol. 12685, 2017.

[18] H. Cheng, J. Liu, T. Xu, B. Ren, J. Mao, and W. Zhang, “Machine learning based low-rate DDoS attack detection for SDN enabled IoT networks,” International Journal of Sensor Networks, vol. 34, no. 1, pp. 56-69, 2020.

[19] S. Velliangiri, P. Karthikeyan, and V. Vinoth Kumar, “Detection of distributed denial of service attack in cloud computing using the optimization-based deep networks,” Journal of Experimental & Theoretical Artificial Intelligence, vol. 33, no. 3, pp. 405-424, 2021.

[20] T. Sindia and J. P. M. Dhas, “A Bifold Software Defined Networking based Defence Mechanism for DDOS Attacks in the Cloud Environment,” International Journal of Applied Engineering Research, vol. 12, no. 20, pp. 9467-9474, 2017.

[21] M. M. Oo, S. Kamolphiwong, T. Kamolphiwong, and S. Vasupongayya, “Analysis of Features Dataset for DDoS Detection by using ASVM Method on Software Defined Networking,” International Journal of Networked and Distributed Computing, vol. 8, no. 2, pp. 86-93, 2020.

[22] O. Hannache and M. C. Batouche, “Neural network-based approach for detection and mitigation of DDoS attacks in SDN environments,” International Journal of Information Security and Privacy (IJISP), vol. 14, no. 3, pp. 50-71, 2020.

[23] K. Sadiq, A. Thompson, and O. Ayeni, “Mitigating DDoS Attacks in Cloud Network using Fog and SDN: A Conceptual Security Framework,” 2020.

[24] M. Zareapoor, P. Shamsolmoali, and M. A. Alam, “Advance DDOS detection and mitigation technique for securing cloud,” International Journal of Computational Science and Engineering, vol. 16, no. 3, pp. 303-310, 2018.

[25] M. A. Aladaileh, M. Anbar, I. H. Hasbullah, Y.-W. Chong, and Y. K. Sanjalawe, “Detection techniques of distributed denial of service attacks on software-defined networking controller–a review,” IEEE Access, vol. 8, pp. 143985-143995, 2020.

[26] N. N. Tuan, P. H. Hung, N. D. Nghia, N. V. Tho, T. V. Phan, and N. H. Thanh, “A DDoS attack mitigation scheme in ISP networks using machine learning based on SDN,” Electronics, vol. 9, no. 3, p. 413, 2020.

[27] Ö. Tonkal, H. Polat, E. Başaran, Z. Cömert, and R. Kocaoğlu, “Machine Learning Approach Equipped with Neighbourhood Component Analysis for DDoS Attack Detection in Software-Defined Networking,” Electronics, vol. 10, no. 11, p. 1227, 2021.

[28] N. Z. Bawany, J. A. Shamsi, and K. Salah, “DDoS attack detection and mitigation using SDN: methods, practices, and solutions,” Arabian Journal for Science and Engineering, vol. 42, no. 2, pp. 425-441, 2017.

[29] 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.

[30] B. Wang, Y. Zheng, W. Lou, and Y. T. Hou, “DDoS attack protection in the era of cloud computing and software-defined networking,” Computer Networks, vol. 81, pp. 308-319, 2015.

[31] T. A. Tuan, H. V. Long, L. H. Son, R. Kumar, I. Priyadarshini, and N. T. K. Son, “Performance evaluation of Botnet DDoS attack detection using machine learning,” Evolutionary Intelligence, vol. 13, no. 2, pp. 283-294, 2020.

[32] T. V. Phan and M. Park, “Efficient distributed denial-of-service attack defense in SDN-based cloud,” IEEE Access, vol. 7, pp. 18701-18714, 2019.

[33] I. Tsamardinos, G. Borboudakis, P. Katsogridakis, P. Pratikakis, and V. Christophides, “A greedy feature selection algorithm for Big Data of high dimensionality,” Machine learning, vol. 108, no. 2, pp. 149-202, 2019.

[34] B. Charbuty and A. Abdulazeez, “Classification based on decision tree algorithm for machine learning,” Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20-28, 2021.

[35] Architecture and Technology (KAT/ACT), D.o.U.o.B.C. (EHU/UPV), gureKddcup database, 2019.

Mono Camera-based Human Skeletal Tracking for Squat Exercise Abnormality Detection using Double Exponential Smoothing

Muhammad Nafis Hisham[1], Mohd Fadzil Abu Hassan[2]*, Norazlin Ibrahim[3], Zalhan Mohd Zin[4]
Universiti Kuala Lumpur, Malaysia France Institute, Bangi, Selangor, Malaysia[1, 2, 3, 4]
UniKL Robotics and Industrial Automation Center (URIAC), UniKL MFI, Bangi, Selangor, Malaysia[2, 3, 4]

Abstract—**Human action analysis is an enthralling area of research in artificial intelligence, as it may be used to improve a range of applications, including sports coaching, rehabilitation, and monitoring. By forecasting the body's vital position of posture, human action analysis may be performed. Human body tracking and action recognition are the two primary components of video-based human action analysis. We present an efficient human tracking model for squat exercises using the open-source MediaPipe technology. The human posture detection model is used to detect and track the vital body joints within the human topology. A series of critical body joint motions are being observed and analysed for aberrant body movement patterns while conducting squat workouts. The model is validated using a squat dataset collected from ten healthy people of varying genders and physiques. The incoming data from the model is filtered using the double exponential smoothing method;the Mean Squared Error between the measured and smoothed angles is determined to classify the movement as normal or abnormal. Level smoothing and trend control have parameters of 0.8928 and 0.77256, respectively. Six out of ten subjects in the trial were precisely predicted by the model. The mean square error of the signals obtained under normal and abnormal squat settings is 56.3197 and 29.7857, respectively. Thus, by utilising a simple threshold method, the low-cost camera-based squat movement condition detection model was able to detect the abnormality of the workout movement.**

Keywords—Abnormality movement; double exponential smoothing; skeletal tracking; mediapipe; squat exercise

I. Introduction

Human activity recognition is a critical application in the computer vision community. Human activity recognition is comprised of two primary components: body tracking and action recognition [1]. These two components have garnered considerable attention in recent years as a result of their multiple applications in areas such as health tracking, sign language recognition, and video surveillance.

Human body tracking can be utilised to tackle a range of problems. This includes avoiding injury during physical body exercise routines (aerobic, anaerobic, or agility training) by monitoring and predicting the person's body's vital points through each frame of a video stream [2]. Injury-free is essential during physical exercise. Thus, computer-assisted self-training systems for sports and exercise can help participants improve their performance and avoid injuries [3]. Besides human body monitoring can be used to solve several issues. Injuries can be avoided during physical training (aerobic, anaerobic, or agility training) by monitoring and predicting body vital points through video streams [2]. Injury-free exercising is crucial and computer-assisted self-training systems for sports and exercise can help athletes increase their performance while avoiding injury [3]. Determining the fundamental cause of postural, balance, and total body coordination difficulties can also be done through corrective exercise.

Action recognition can also benefit the living. During the COVID-19 pandemic, Malaysians were advised to stay at home and avoid routine visits. This may harm older people who were not accompanied by youngsters. Due to health difficulties, most elderly people are in danger of falling and fainting. Seniors may also face security issues such as robbery. Moreover, a study found that Malaysian healthcare has improved, accelerating the "Silver tsunami" of population ageing [4]. Thus, eldercare should focus on this population. Using machine learning and contemporary computer vision techniques, we can detect things of interest more accurately than humans [5]. Hence, an automated eldercare system should be considered for good monitoring.

Additionally, numerous studies have demonstrated that commercially available devices such as the Microsoft Kinect Sensor, the PlayStation Eye, and the Wiimote are successful at sensing and analysing human joint motion in many applications [6]. However, these researcher tracking studies necessitated the use of a 3-Dimensional (3D) sensor and additional expensive visual and wearable sensors. As a result, a low-cost mono camera equipped with a MediaPipe algorithm is proposed for determining pose landmarks and monitoring action motion in daily routine human activity.

II. Literature Review

Human skeletal tracking consists of tracking several key points from different parts of the human body, such as the body joints, eyes, nose, and ears depending on the purpose of tracking. The key points are connected creating a human skeletal form. Generally, the tracking of human motion can be done through visual information like images from video or extraction data through camera sensors. From the human skeletal features, Human Pose Estimation (HPE) and Human Action Recognition (HAR) can be applied. HPE is widely

*Corresponding Author.
Short Term Research Grant (Str19023) UniKL

applied to solve various problems such as bad posture correction, sports movement monitoring, fall detection, etc., [7]. HAR can be used to recognize daily life activities such as running, jumping, squatting, etc. which are recognized from video sequences using the vectors extracted [8]. The goal of HAR is to study and recognize the nature of action from unknown video sequences automatically [9].

A. Human Body and Movement Tracking

Human movement detection predicts and tracks an entity's position and orientation within an image/video frame. Previous researchers have investigated various techniques for detecting human movement. Certain detecting methods make use of wearable sensors. The Inertial Motion Unit (IMU) sensor is one of the most frequently used wearable sensors [10]. IMU technology is used in conjunction with a machine learning (ML) technique to precisely locate body key points and deliver accurate orientation measurements. However, relying exclusively on wearable sensors may impose significant limits, particularly for home-based monitoring. In addition, the implementation of the environmental sensor leads to high costs due to the use of professional external sensors [11].

In comparison to prior systems, some researchers use depth cameras such as the Kinect sensor, a low-cost consumer-grade 3D camera, to extract key points on the human body [12]. This technology can contribute to the provision of several human action characteristics, such as depth, colour differentiation, and human skeletal structure [13]. The x and y coordinates, as well as the confidence value, were recorded for each body key point, and a 2-Dimensional (2D) human movement estimate can be performed using the OpenPose algorithm [14]. Human movement detection can also be performed with an RGB stereo camera as a result of enhanced advances in in-depth imaging technology [6].

B. Machine Learning Framework

A framework represents an interface algorithm that makes the work of the machine learning model easier and faster. [7] apply six machine learning frameworks; OpenPose, Tflite, Pifpat, Tfjs (mobileNet), Tfjs (Resnet 50), and BlazePose for pose estimation and correction of fitness training dataset since the goal was to find the fastest and most accurate methods that work in real-time. Besides that, the use of a Kinect sensor to extract the skeletal features has also been used to determine the abnormality of squat exercise from a dataset [12]. In addition, [15] and [16] suggested the use of the OpenPose framework to extract the keypoints coordinates from the RGB data for pose detection of fall action. In comparison, the BlazePose framework that was used by the MediaPipe model presented a topology with 33 human body keypoints, which is more than OpenPose and Kinect topologies, which only provide 17 human body keypoints [2].

MediaPipe is an open-source ML model purposely for live and streaming media and provides direct and customizable Python solutions as a prebuilt Python package [17]. The primary use of MediaPipe is the quick creation of perceptual pipelines with any models and other reusable components [18]. MediaPipe offers several solutions such as MediaPipe Face Detection, MediaPipe Hands, MediaPipe Pose, MediaPipe Object Detection, etc. MediaPipe Hands tracking solution offers an ML pipeline that includes two different models that act in tandem; Palm Detection Model and Hand Landmark Model [19].

Furthermore, The MediaPipe Pose model, which was inspired by Leonardo's Vitruvian man, can identify the human body by predicting the midpoint of the human's hip, the radius of a circle circumscribing the full body, and the inclination angle of the line linking the shoulder and hip midpoints [18]. In [7] the author applied the MediaPipe Pose solution, which can predict 33 3D human body landmarks for human pose assessment and correction with an emphasis on fitness training. The MediaPipe Pose solution was also used during the pre-trained pose estimate for data collection of various yoga poses [20].

C. Rehabilitation Exercise

Physical therapy and rehabilitation programs are critical for persons who participate in sports, are accident victims, or are senior citizens. Nowadays, rehabilitation programs are conducted prior to the occurrence of any sports injury in order to prepare the athletes for the next level of physical demands [21]. The rehabilitation approach will involve exercises to increase muscle and joint range of motion. To ensure a consistent recovery outcome, patients are urged to complete a series of prescribed home-based physiotherapy sessions.

Squats are a well-known rehabilitation exercise, particularly for people with knee difficulties or injuries. This exercise is ideal for a home workout program because it does not require any extra equipment. Squat exercises require the patient's knee and hip joints to flex and extend to develop the body part's flexion and extension strength [22]. Hip, knee, and ankle movements must be synchronized correctly to ensure efficient muscle function and avoid damage throughout the activity [23]. The squat exercise cycle is performed as followed:

Step 1: Stand straight, open your feet align with the shoulder, and stretch out your arms to the front;

Step 2: Bend down your knees slowly until you reach a half-crouched position while maintaining your chest upward and back straight;

Step 3: Hold the previous step position for approximately one second while maintaining your feet flat on the floor;

Step 4: Slowly stand up to the initial position and one cycle is completed;

Step 5: Repeat steps 1 to 4.

All of the procedures above must be followed precisely to guarantee that the exercise's objective is met and that undesired injury is avoided. The training and testing phases of this paper make use of a similar squat dataset from [14]. The dataset was created by video recording squat exercises performed by ten healthy volunteers of varying gender and body build using Microsoft Kinect. The dataset is divided into two distinct categories: normal and aberrant. The normal condition dataset contains somewhat slow squatting movements to mimic the patient's gradual movement during rehabilitation. While the abnormal dataset was conducted with the same group of volunteers performing a comparable squat exercise with only one leg. This is done to mimic the same level of discomfort experienced by some patients during therapy.

III. Methodology

The goal of this research is to identify the abnormality of squat exercise by a single human using a single camera image. Training and testing datasets for classifiers were acquired from a set of ten subjects with different genders and physical body postures performed half-squat exercise. The diversities of human squat exercise were classified into two categories; normal condition squat and abnormal condition squat as shown in Fig. 1.

The normal condition squat consists of ten videos of the different subjects performing double legs half-squat exercise as shown in Fig. 1(a). On the other hand, the abnormal condition squat videos consist of ten videos of a similar subject in normal condition squat performing single legs half-squat exercise as shown in Fig. 1(b). Each subject conducted the exercise separately, and it was captured from front viewpoint angles in normal lab lighting. In this study, the distance between the 3D image sensor and subjects is fixed in the range of 2.5-3 meters.

A. *Human Skeletal Tracking-MediaPipe Pose*

MediaPipe Pose is the current solution for human pose assessment i.e. fitness training [12] which can predict 33 3D human body landmarks as shown in Fig. 2. By using the MediaPipe Pose model, this study is improvised and able to track the subjects' lower body (from hip to ankle) of the squat exercise dataset without the use of any marker on 2D images.

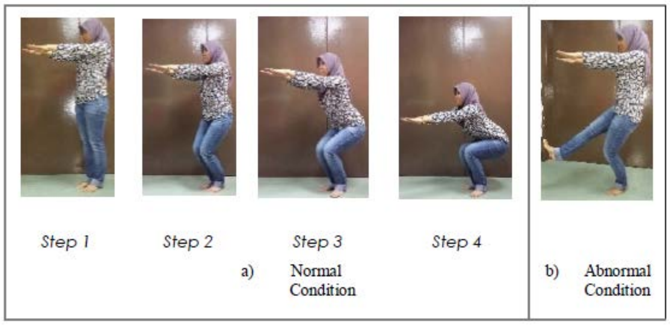


Fig. 1. Normal and Abnormal Squat Exercise Posture Steps [10].

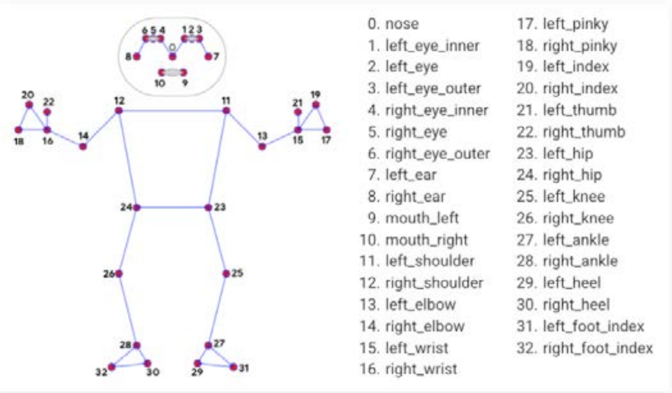


Fig. 2. Thirty-Three Human Skeletal Landmarks of MediaPipe Pose Model.

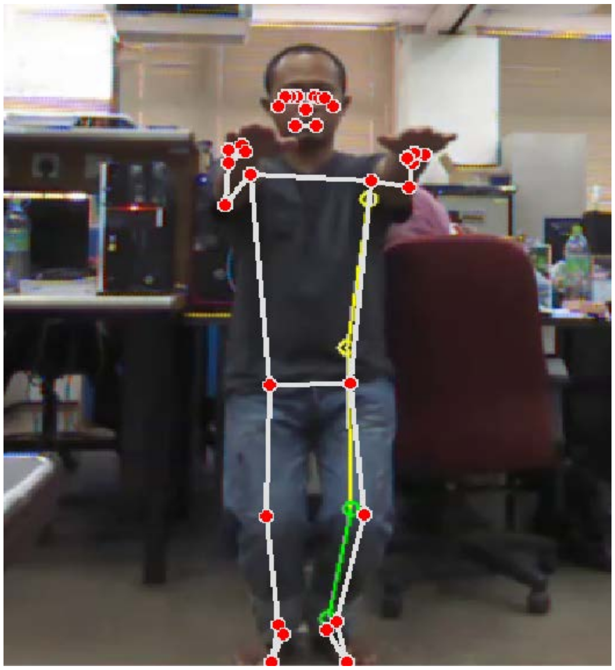


Fig. 3. Thirty-Three 3D Body Landmarks-MediaPipe Pose on 2D Image.

The study focuses on the hip-knee-ankle of a subject performing the half-squat exercise. The MediaPipe Pose model is capable to predict the essential human body landmarks as shown in Fig. 3.

B. *Joint Angle Tracking*

The angle was generated in 3D space based on the joint's coordinates as shown in Fig. 4. Euclidean distance is computed to evaluate the distance between two joints; hip-knee joint, knee-ankle joint, and knee-hip joint [24] using (1)-(3) below:

$$d_{HL,KL} = \sqrt{(x_{HL}-x_{KL})^2+(y_{HL}-y_{KL})^2+(z_{HL}-z_{KL})^2} \quad (1)$$

$$d_{KL,AL} = \sqrt{(x_{KL}-x_{AL})^2+(y_{KL}-y_{AL})^2+(z_{KL}-z_{AL})^2} \quad (2)$$

$$d_{AL,HL} = \sqrt{(x_{AL}-x_{HL})^2+(y_{AL}-y_{HL})^2+(z_{AL}-z_{HL})^2} \quad (3)$$

From the distance calculated, the knee angle (θ), is determined by using the Law of Cosines in (4) as stated below:

$$\theta_K = \cos^{-1}\left(\frac{d_{HL,KL}^2 + d_{KL,AL}^2 - d_{AL,HL}^2}{2d_{HL,KL}d_{KL,AL}}\right), 0 \leq \theta_K \leq \pi \quad (4)$$

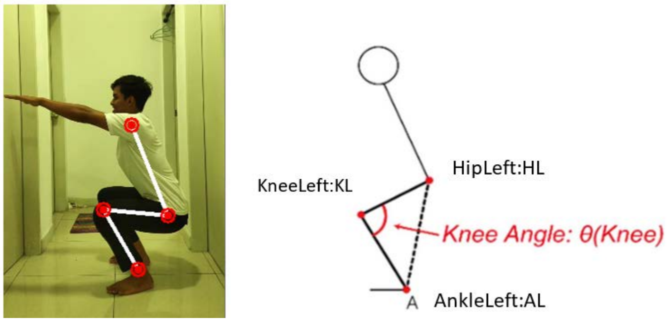


Fig. 4. Illustration of Tracking Lower Body Joints and Joint Movement.

C. Signal Filtering

From the result of the hip-knee-ankle angle extracted from the skeletal features, it is discovered that there was some unwanted noise presented. This random noise might occur due to the image parameter such as illumination disturbance, random shadow, subject pose, etc. from the dataset [12]. Therefore, an algorithm is applied to filter the incoming data image for the elimination of the noise before proceeding to the next process.

Double Exponential Smoothing (DES) was chosen for the following task as suggested by [25] to predict the trend for forecasting. Generally, DES is approached in economic data analysis to predict the immense range of noise, to forecast a trend, and give high forecasting accuracy level. DES also has been proven to run 100 times faster than the extended Kalman Filter [26]. The (5)-(6) are applied as stated below which are related to DES:

$$S_t = \alpha y_t + (1-\alpha)(S_{t-1} + b_{t-1})\,,\alpha\ \in [0{,}1] \qquad (5)$$

$$S_t = \alpha y_t + (1-\alpha)(S_{t-1} + b_{t-1})\,,\alpha\ \in [0{,}1] \qquad (6)$$

where:

y_t = original measured angle.

α = level smoothing parameter.

S_t = smoothed value.

b_t = trend control parameter.

β = trend smoothing parameter.

Equation 5 is applied for adjusting the new smoothed value (S_t) by totaling the previous smoothed value (S_t-1) with the previous trend control (b_t-1). Equation 6 is responsible for updating the new trend control (b_t) by calculating the difference between current S_t with S_t-1. In this study, the initial value of S_t and b_t are $S_1 = y_1$ and b1 = $y_2 - y_1$; similar to the method applied by [10]. All the angles' data of the 10 samples are recorded and smoothed using DES. Fig. 5 represents one of the examples of DES application on the original angle data and the result of the smoothed value. It shows that smoothed using DES is able to reduce the random noise due to environmental constraints.

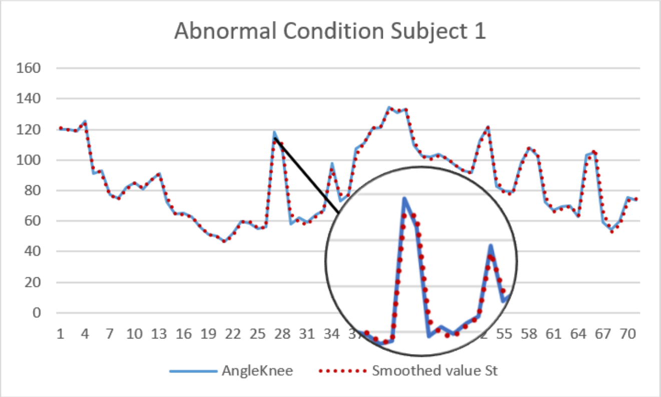


Fig. 5. Original Angle Vs Smoothed Angle Data.

Next, the mean squared error (MSE) is calculated from the smoothed value of the knee angle by the DES method [25]. The MSE method is used to determine the values of constant that minimizes the error size and how close estimations or forecasts are to actual data. The MSE formula can be defined in (7) below:

$$\mathrm{MSE} = \frac{1}{n}\sum e_t^2 \qquad (7)$$

Where the error, e, is derived from the difference of values between the current period of the original measured angle (yt) and the previous period of smoothed value (S_{t-1}) as shown in (8).

$$e_t = y_t - S_{t-1} \qquad (8)$$

In this study, to calculate the MSE value for all samples of the normal and abnormal conditions, the average value is calculated. The optimum α and β values of 0.8928 and 0.77256 respectively are reused based on performance results obtained by [10].

IV. Result

Based on the experimental setup explained in Section II, Fig. 6(a) shows the half-squat exercise in normal conditions, while Fig. 6(b) shows half-squat exercise in abnormal condition performed by subject 1. The hip-knee-ankle angle was extracted from the squat exercise performed by subject 1 in normal conditions (Fig. 7) and abnormal conditions (Fig. 8).

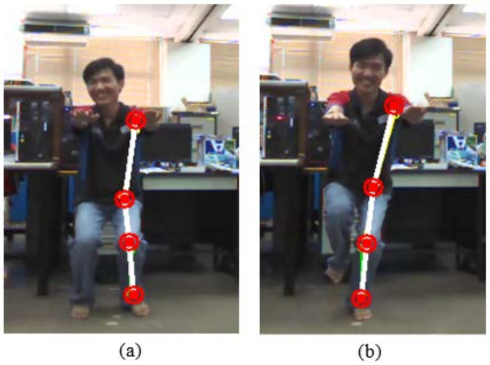


Fig. 6. Half-Squat Exercise in Normal (a) and Abnormal (b) Conditions.

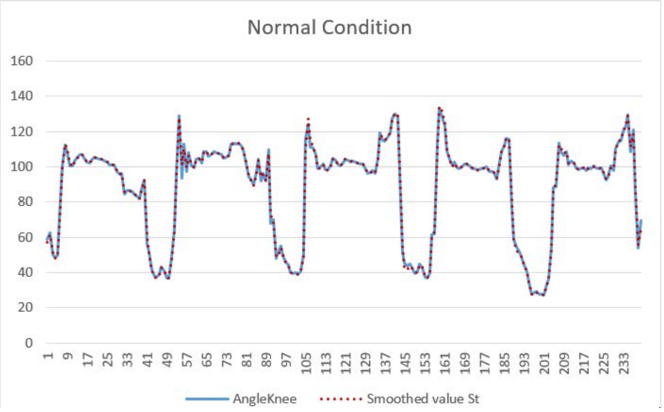


Fig. 7. Knee Angle Captured from Normal Squat Exercise for Subject 1.

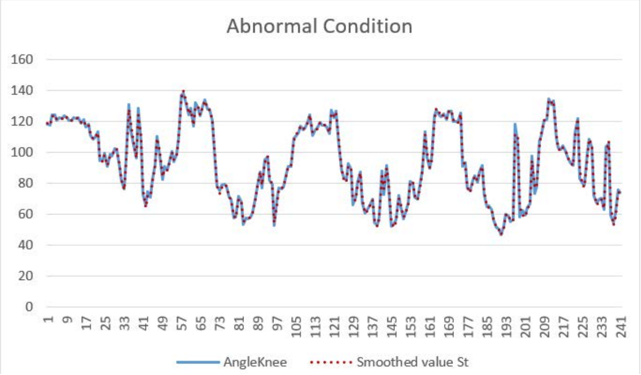


Fig. 8. Knee Angle Captured from Abnormal Squat Exercise for Subject 1.

Based on the data that have been collected from the dataset in Table I, there is a significant difference between the result of MSE of the normal and abnormal conditions. The average of MSE from 10 normal exercise data is 50.63, for the specified constants: α = 0.8928 and β = 0.7726. In contrast, the average MSE for ten abnormal exercise data sets is 84.69.

The comparison of MSE normal and abnormal conditions, the difference between MSE, and the average of MSE among the subjects are presented in Fig. 9. From the figure, subjects 4 and 8 performed adequate accurate movements during both normal and abnormal squat exercises in comparison to the other subjects. This shows that their positions are balanced during both normal and abnormal exercise. In contrast, squat exercise movement of subjects 5 and 10 can be considered mediocre based on the difference between the MSE of normal and abnormal. The threshold value of 67.66 (mean MSE) meets as a dividing boundary between the normal and abnormal conditions for all subjects except subjects 1, 4, 6, and 8.

TABLE I. THE PERFORMANCE RESULT FOR EACH SUBJECT

Subject No.	**Mean Squared Error**			
	Normal	***Abnormal***	***Difference***	***Average***
1	116.40	151.83	35.43	134.11
2	40.94	86.45	45.51	63.69
3	43.25	92.61	49.36	67.93
4	50.13	62.12	11.99	56.12
Subject No.	**Mean Squared Error**			
	Normal	***Abnormal***	***Difference***	***Average***
5	54.18	106.29	52.16	80.24
6	27.14	58.48	31.33	42.81
7	49.73	71.50	21.77	60.62
8	39.16	52.25	13.08	45.71
9	50.42	80.33	29.91	65.38
10	34.95	85.04	50.09	59.99
Mean	**50.63**	**84.69**	**34.06**	**67.66**
Min	**27.14**	**52.25**	**11.99**	**42.81**
Max	**116.40**	**151.83**	**52.12**	**134.11**

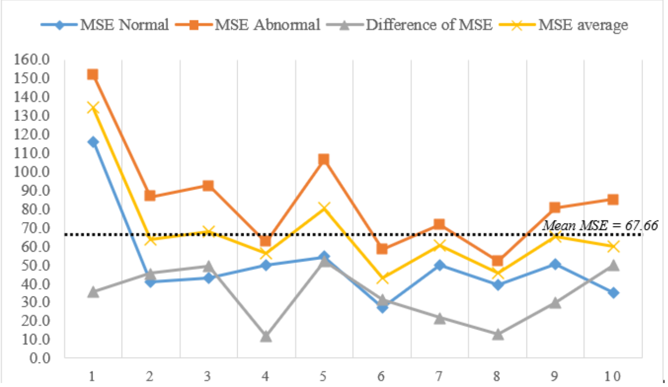


Fig. 9. MSE for Normal Vs Abnormal Squat Condition.

V. Discussion

From Fig. 7 and Fig. 8, the trend in MSE of the normal condition has a lower frequency of vibration compared with the abnormal condition. This showed that the subject was able to maintain a better squat form while performing the exercise in normal conditions compared to an abnormal condition. This proves that MSE is a suitable measurement for evaluating the performance of the system [27]. The smoothed value also helps in evaluating the data by reducing the noise. Thus, the simulated result shows that the DES method has been proven as one of the methods to reduce random noise from the original data of the dataset based on the computed results [28]. As a result, from Table I, the average MSE for normal condition squat exercise is lower and better than the abnormal condition. This shows that the MediaPipe Pose model able to predict the 3D body landmarks of subjects efficiently [12].

Fig. 9 shows that their positions are balanced during both normal and abnormal exercise. In contrast, squat exercise movement of subjects 5 and 10 can be considered mediocre based on the difference between the MSE of normal and abnormal. The threshold value of 67.66 (mean MSE) meets as a dividing boundary between the normal and abnormal conditions for all subjects except subjects 1, 4, 6, and 8.

There are some limitations of the model observed in this research that need to be considered and further addressed. Firstly, beside its high ability to perform human detection of squat action, MediaPipe Pose algorithm has so far only managed to detect a single person per frame. It may not be not suitable to be used for human detection of squat action when multiple humans present in the image frames. In addition to that, the model faces difficulties if there are obstacles blocking the camera's view of human upper body. This will affect the human detection confidence level. To address this issue and to have better detection performance, it is advisable to ensure that no foreign object block the camera view. Lastly, like other vision system issues, the performance of human detection might be affected due to the lack of proper illumination.

VI. Conclusion

This paper introduced the use of the MediaPipe Pose model for human movement tracking on a 2D frame-frame video dataset. The MediaPipe Pose model is a powerful model which can predict thirty-three 3D human body landmarks and can be used to perform a human virtual skeletal model for features extraction. Skeletal pose features will contain all of the necessary information to comprehend the action's results. This model can be applied to a rehabilitation center or self-monitoring exercise at home. DES method was able to minimize the random noise from the tracked angle data that might be due to the image parameter such as illumination, shadow, pose, etc. Next, MSE is applied to evaluate the movement cycle for squat exercise. Finally, the mean MSE was used as the threshold value to differentiate the posture movement for squat exercise between the two conditions.

ACKNOWLEDGMENT

The authors would like to express their gratitude to UniKL MFI for giving support in conducting this research and the financial support from Short Term Research Grant (Str19023) UniKL.

REFERENCES

[1] Luvizon, D. C., Picard, D., & Tabia, H., "2D/3D Pose Estimation and Action Recognition Using Multitask Deep Learning", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018.

[2] Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M., "BlazePose: On-device Real-time Body Pose Tracking", 2020.

[3] Yadav, S. K., Singh, A., Gupta, A., & Raheja, J. L., "Real-time Yoga Recognition using Deep Learning", Neural Computing and Applications, 2019.

[4] Abu Hassan, M. F. , Hussain, A., Muhamad, M. H., & Yusof, Y., "Convolution Neural Network-based Action Recognition for Fall Event Detection", International Journal of Advanced Trends in Computer Science and Engineering, 2019.

[5] Cameron, J. A., Savoie, P., Kaye, M. E., & Scheme, E. J., "Design Considerations for the Processing System of a CNN-based Automated Surveillance System", Expert Systems with Applications, 2019.

[6] Abu Hassan, M. F., Hussain, A., Md Saad, M. H., & Win, K., "3D Distance Measurement Accuracy on Low-cost Stereo Camera", Science International Journal, 2017.

[7] Ohri, A., Agrawal, S., & Chaudhary, G. S., "On-device Realtime Pose Estimation & Correction", International Journal of Advances in Engineering and Management (IJAEM), 2021.

[8] Abdellaoui, M., & Douik, A. (2020). Human Action Recognition in Video Sequences Using Deep Belief Networks. Traitement du Signal, 37(1).

[9] Badiola-Bengoa, A., & Mendez-Zorrilla, A. (2021). A Systematic Review of the Application of Camera-Based Human Pose Estimation in the Field of Sport and Physical Exercise. Sensors, 21(18), 5996. doi:10.3390/s21185996.

[10] Wilk, M. P., Walsh, M., & O'Flynn, B., "Multimodal Sensor Fusion for Low-Power Wearable Human Motion Tracking Systems in Sports Applications", IEEE Sensors Journal, 2020.

[11] Han, K., Yang, Q., & Huang, Z. (2020). A two-stage fall recognition algorithm based on human posture features. Sensors, 20(23), 6966.

[12] Abu Hassan, M. F., Zulkifley, M. A., & Hussain, A., "Squat Exercise Abnormality Detection by Analyzing Joint Angle for Knee Osteoarthritis Rehabilitation", Jurnal Teknologi, 2015.

[13] Tasnim, N., Islam, M., & Baek, J. H., "Deep Learning-based Action Recognition using 3D Skeleton Joints Information", Inventions, 2020.

[14] Wu, Q., Xu, G., Zhang, S., Li, Y., & Wei, F., "Human 3D Pose Estimation in a Lying Position by RGB-D Images for Medical Diagnosis and Rehabilitation", 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020.

[15] Huang, Z., Liu, Y., Fang, Y., & Horn, B. K. (2018, October). Video-based fall detection for seniors with human pose estimation. In 2018 4th International Conference on Universal Village (UV) (pp. 1-4). IEEE.

[16] Wang, B. H., Yu, J., Wang, K., Bao, X. Y., & Mao, K. M. (2020). Fall detection based on dual-channel feature integration. IEEE Access, 8.

[17] MediaPipe, " Live ML Anywhere", Accessed on: March 3, 2022. [Online] Available: https://mediapipe.dev/.

[18] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.

[19] Halder, A., & Tayade, A. (2021). Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning. International Journal of Research Publication and Reviews, 2.

[20] Anilkumar, A., KT, A., Sajan, S., & KA, S. (2021). Pose Estimated Yoga Monitoring System. Available at SSRN 3882498.

[21] Tang, D., "Hybridized Hierarchical Deep Convolutional Neural Network for Sports Rehabilitation Exercises", in IEEE Access, vol. 8, 2020.

[22] Akagi, R., Sato, S., Hirata, N., Imaizumi, N., Tanimoto, H., Ando, R., & Hirata, K., "Eight-week Low-intensity Squat Training at Slow Speed Simultaneously Improves Knee and Hip Flexion and Extension Strength", Frontiers in Physiology, 2020.

[23] Ribeiro, A. S., Santos, E. D., Nunes, J. P., Nascimento, M. A., Graça, Á., Bezerra, E. S., & Mayhew, J. L. (2022). A Brief Review on the Effects of the Squat Exercise on Lower-Limb Muscle Hypertrophy. Strength & Conditioning Journal.

[24] Islam, M. S., Bakhat, K., Khan, R., Iqbal, M., Islam, M. M., & Ye, Z., "Action Recognition using Interrelationships of 3D Joints and Frames based on Angle Sine Relation and Distance Features using Interrelationships", Applied Intelligence, 2021.

[25] Alhindawi, R., Abu Nahleh, Y., Kumar, A., & Shiwakoti, N., "Projection of Greenhouse Gas Emissions for the Road Transport Sector based on Multivariate Regression and the Double Exponential Smoothing Model", Sustainability, 2020.

[26] Chung, M. G. & Kim, S. K., "Efficient Jitter Compensation using Double Exponential Smoothing", Information Sciences, 2013.

[27] Hodson, T. O., Over, T. M., & Foks, S. S. (2021). Mean squared error, deconstructed. Journal of Advances in Modeling Earth Systems, 13(12), e2021MS002681.

[28] Nazim, A., & Afthanorhan, A. (2014). A comparison between single exponential smoothing (SES), double exponential smoothing (DES), holt's (brown) and adaptive response rate exponential smoothing (ARRES) techniques in forecasting Malaysia population. Global Journal of Mathematical Analysis, 2(4), 276-280.

Solving the Imbalanced and Limited Data Labeled for Automated Essay Scoring using Cost Sensitive XGBoost and Pseudo-Labeling

Marvina Pramularsih[1], Mardhani Riasetiawan[2]*
Department of Computer Science and Electronics
Faculty of Mathematics and Natural Sciences
Universitas Gadjah Mada, Yogyakarta, Indonesia

***Abstract*—There are two main problems on forming the Automatic Essay Scoring Model. They are the datasets having imbalanced amount of the right and wrong answers and the minimal use of labeled data in the model training. The model forming based on these problems is divided into three main points, namely word representation, Cost-Sensitive XGBoost Classification, and adding unlabeled data with the Pseudo-Labeling Technique. The essay answer data is converted into a vector using the trained word vector fastText. Furthermore, the classification of unlabeled data was carried out using the Cost-Sensitive XGBoost Method. The data labeled by the classification model is added as training data for the new classification model form. The process is carried out iteratively. This research is about using the combination of Cost-Sensitive XGBoost Classification and Pseudo-Labeling which is expected to solve the problems. For the 0th iteration, the dataset having a ratio of the amount of "right" labeled data with the amount of "right" labeled data is close to 1, in other words a balanced dataset or a ratio that is more than 1 produces a model with better performance. Thus, the selection of training data at an early stage must pay attention to this ratio. In addition, the use of the Hybrid Method on these datasets can save labeled data 56 times compared to the AdaBoost Method. Hybrid model is able to produce F1-Measure more than 95.6%, so it can be concluded that the Hybrid Method, which combines the XGBoost and Pseudo-Labeling Cost-Sensitive Classification with Self Training, is able to overcome the problem of unbalanced datasets and data limited label.**

***Keywords*—Imbalanced data; limited labeled data; automated essay scoring; cost sensitive XGBoost; pseudo-labeling**

I. Introduction

Pusat Asesmen dan Pembelajaran (PUSMENJAR), Ministry of Cultural, Education, and Research Technology, Republic of Indonesia conducts a mapping program of educational attainment to monitor the quality of education nationally or locally, called AKSI (Asesmen Kompetensi Siswa Indonesia). Learning evaluation is carried out based on the results of the questions tested on students. PUSMENJAR distinguishes question packages into two, those are literacy and numeracy. There are eight packages of literacy questions and eight packages of numeracy questions that will be used to simulate the exam. In the sixteen question packages, there are six types of questions, namely true or false, check box, matching, sorting, multiple choice, and essay. For the process of correcting true or false, check boxes, matching, sorting, and multiple choice can be done by matching the answer keys. However, the process of correcting essay answers cannot always be done by matching an answer with the answer key. In addition, essay answers scoring manually takes longer than answers for multiple choice questions and short answers [1]. Therefore, we need models for scoring essay answers automatically.

Herwanto et al. [2] sees correction of the essay answer as a classification of a true or false answer using AdaBoost Classification Method. Another Automated Essay Scoring (AES) model has begun to be researched to develop a model for automatically correcting essay answers in Indonesian [3]. However, the models that have been developed have not paid attention to the effect of the large number of manually labeled data on model performance and have not paid attention to whether the dataset is a dataset that has amount of the right and wrong labeled data and is balanced or not.

The exam simulation of high school level questions obtained ten datasets of essay answers. After manual labeling by experts, it is known that the dataset is an imbalanced dataset between correct and false answers. Therefore, a classification algorithm which is capable of handling imbalanced data characteristics is needed. Fernandez et al. [4] and He and Ma [5] state that the approach taken to solve imbalanced data problems is divided into three points, namely methods at the data level, methods at the algorithm level, and methods at the hybrid level. The method at the algorithm level is the easiest method to apply [6]. Wang et al. [7] and Xia et al. [6] classify imbalanced data using the Cost-Sensitive XGBoost method. The use of the XGBoost algorithm is due to the fact there are many teams wining the competition using this algorithm [8]. In addition to the imbalance problem in the dataset, there is another problem that is forming a model with good performance with less training data than the current number of labeled data. The number of labeled data provided is currently around 6,000 data and will be used to correct approximately 330,000 uncorrected answer data. This problem can be overcome, one of which is by implementing Pseudo-Labeling with Self Training as was done by Babakhin et al. [9].

*Corresponding Author.

II. Automated Essay Scoring

Research on Automated Essay Scoring (AES) began since Page [10] conducted research on essay assessment using computers. AES research on answers to essays in Indonesia has begun to be developed from various points of view. One of them views the problem of essay assessment as a problem of classifying right or wrong answers [2], [3].

Herwanto et al. [2] conducted research on the AES Model for answers to essays in Indonesia. The dataset used is three datasets of student answers from the Program for International Student Assessments (PISA). The word representation used is Bag-of-Words (BoW) and character ngrams. The classification algorithm used is Adaptive Boosting (AdaBoost). The AES model formed has a F1-Score of 97.69% for the Machu Picchu dataset, 67.2% for the jacket dataset, and 71.74% for the bicycle dataset. Riasetiawan et al. [3] conducted research for essay answers on clustering and classification. The dataset used is a dataset of essay answers from the Indonesian Ministry of Education and Culture (Kemendikbud). The clustering algorithm used is K-Means, while the classification algorithm used is Convolutional Neural Network (CNN). Prior to clustering and classification, the word representation stage was carried out using GloVe. The answer classification model yields an accuracy above 85%.

In supervised learning, not all labeled datasets are balanced datasets. To solve the problem of unbalanced datasets, it can be done using data sampling methods [11], cost-sensitive algorithms [6], or a combination of data sampling methods and cost-sensitive algorithms [12]. The use of the data sampling method has weaknesses, namely, in addition to choosing a suitable classification method for the dataset, there is a need for further analysis of what data sampling method is more suitable for the dataset before entering the classification model training stage. In a cost-sensitive algorithm, there is no need to add these steps, so this method is the easiest of the other two methods to be applied to an unbalanced dataset.

Xu et al. [11] conducted research on the classification of sentiments and emotions on an unbalanced dataset. By using the Support Vector Machine (SVM) and Word Embedding Compositionality with Minority Oversampling Technique (WEC-MOTE), it can increase the precision by 29.3%. Xia et al. [6] conducted research on the classification of borrowers in peer-to-peer lending. By using the Cost-Sensitive XGBoost (CSXGBoost), the highest Area under the ROC Curve (AUC) value was obtained when compared to all trials compared by Xia et al. [6], which is 74.85%. The use of cost-sensitive is the easiest way to deal with unbalanced datasets [6]. Le et al. [12] conducted research on the classification of bankruptcy companies. Using CBoost and SMOTE-ENN, the highest AUC values were obtained from all the trials compared by Le et al. [12], which is 87.1%. Pseudo-Labeling is a technique used to increase the amount of labeled data by utilizing the unlabeled data that is owned. There are several learning algorithms used in Pseudo-Labeling, namely Self-Training [9], Co-Training [13], and Cluster-then-label [14]. In the Co-Training algorithm, the features used must be divided into two. Meanwhile, in cluster-labeling, there is a clustering stage before labeling. The clustering stage in the Cluster-then-label algorithm has its own challenges, namely making a good cluster. Based on the three algorithms, namely Self-Training [9], Co-Training [13], and Cluster-then-label [14], the Self-Training algorithm is the easiest algorithm to be implemented.

III. Hybrid Method

This research uses eight pairs of datasets, namely eight labeled datasets and eight unlabeled datasets [15], shown in Table I. Each dataset uses the .xlsx format. The datasets are from the PUSMENJAR. The datasets are the answers to the simulation of AKSI questions. For each pair the dataset will be analyzed and the model with the best performance is sought based on the experiments carried out.

Data pre-processing is necessary being used in the next process. The pre-processing consists of Lower-Case Folding, Filtering, and Tokenization, fastText pretrained word embedding for Indonesia. Pre-trained word vectors are trained on Wikipedia using fastText. The model is trained using skip-gram, dimension 100, subwords size 3-6 characters, epoch 5, and learning rate 0.05. To increase the chances that the machine can produce the best labels, the best-performing classification model is needed before it is used to predict answers on the unlabeled dataset. This is done by training the model with several combinations of different parameter values and selecting the model with the highest F1-Measure. Fig. 1 shows the Cost-Sensitive XGBoost modeling development for the research.

TABLE I. DATASET DESCRIPTION

Dataset		**Description**		
	Correct Label	***Wrong Label***	***Unlabeled***	***Total***
A	3.371	2.065	334.301	339.737
B	1.972	*3.824*	237.208	243.004
C	2.298	*3.525*	336.266	342.089
D	1.072	*4.716*	334.082	339.870
E	757	*5.046*	333.094	338.897
F	1.235	*4.605*	333.741	339.581
G	140	*5.566*	333.574	339.280
H	393	*5.288*	332.407	338.088

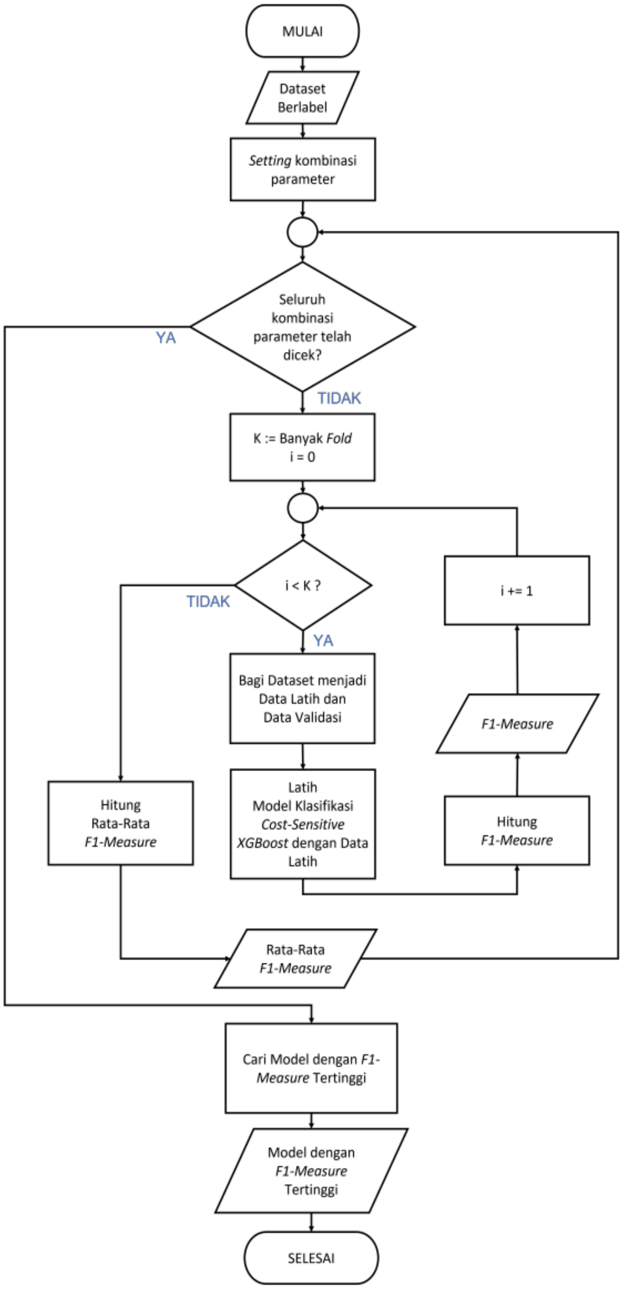

MULAI
Dataset Berlabel
Setting kombinasi parameter
Seluruh kombinasi parameter telah dicek?
YA
TIDAK
K := Banyak Fold
i = 0
i < K ?
TIDAK
YA
i += 1
Bagi Dataset menjadi Data Latih dan Data Validasi
F1-Measure
Hitung Rata-Rata F1-Measure
Latih Model Klasifikasi Cost-Sensitive XGBoost dengan Data Latih
Hitung F1-Measure
Rata-Rata F1-Measure
Cari Model dengan F1-Measure Tertinggi
Model dengan F1-Measure Tertinggi
SELESAI

Fig. 1. Cost-Sensitive XGBoost.

IV. Result and Discussion

Pseudo-Labeling (Fig. 2) testing is seen based on two things, those are the amount of initial data and the amount of data added per iteration. This experiment is repeated until the third iteration. Fig. 3 (a)-(h) are the results of testing the initial data for Pseudo-Labeling. Based on Fig. 3, it is found that in the 0th iteration, the F1-Measure model with N0 = 100 is always lower than the model with more initial data. However, after iterating until the third iteration the F1-Measure of the model increases. On the Dataset 2, 4, 5, and 6, the F1-Measures of the models are able to exceed the models using a bigger amount of manual labeled data. In the Dataset 7, F1-Measure of the model with N0 = 100 is 0%. Although, it has been done for some iterations, but cannot improve the F1-Measure of the model. In this case, the number of “true” labeled data is 1, while the “false” labeled is 99. For the extreme case of the Dataset 7, the Pseudo-Labeling technique has not been able to improve the F1-Measure of the model. Furthermore, the amount of additional data tested are 50, 100, 150, and 200. Fig. 4 (a)-(h) are the results of testing additional data for Pseudo-Labeling using 100 labeled data as initial data. Based on Fig. 4, the selections of the amount of additional data per iteration that have been tried always increase F1-Measure of the models when compared to F1-Measure of the models in the 0th iteration, except for the Dataset 7 which only has 1 “true” labeled data and 99 “false” labeled data. Furthermore, the selection of additional data which is 50 (half of the number of initial data) is always lower than the others (equal to or higher than the initial data).

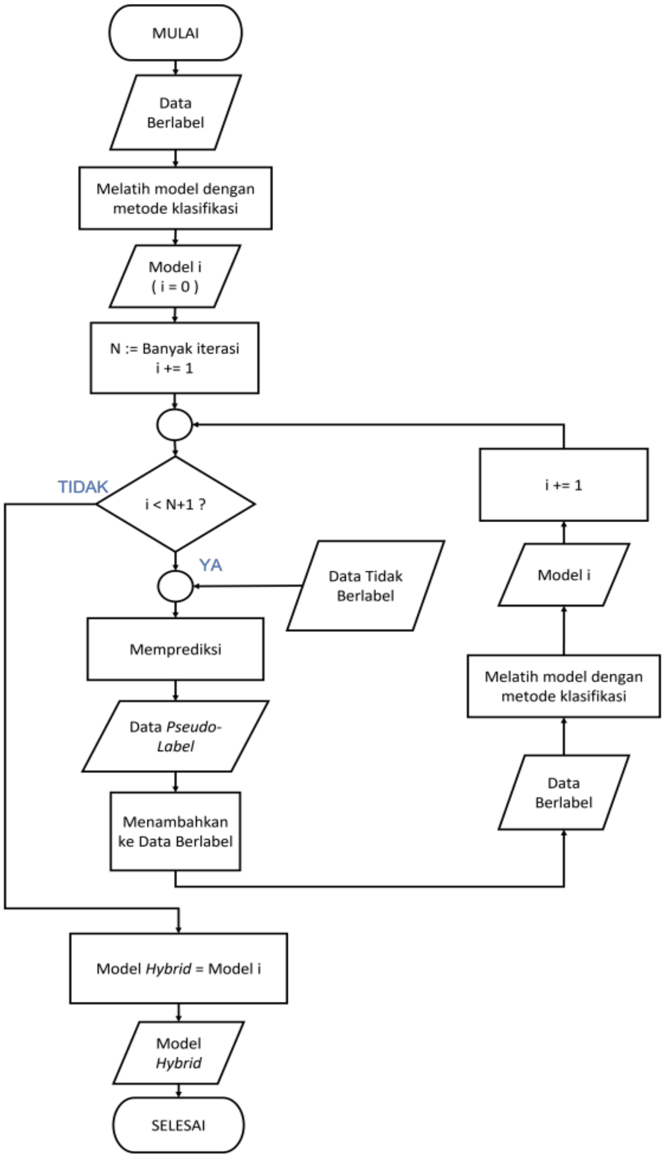
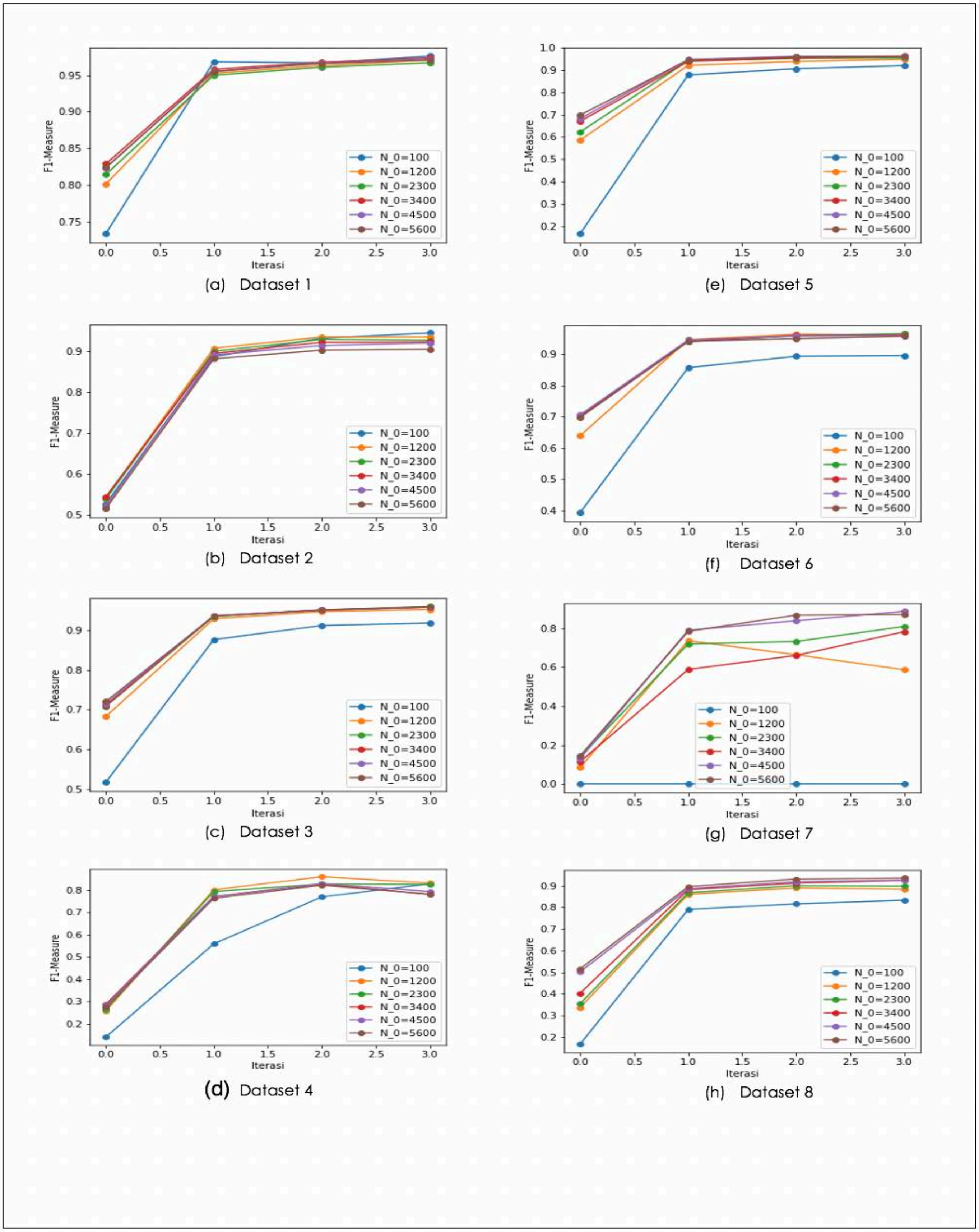


Fig. 2. Pseudo-Labeling with Self-Training.



(a) Dataset 1

(e) Dataset 5

(b) Dataset 2

(f) Dataset 6

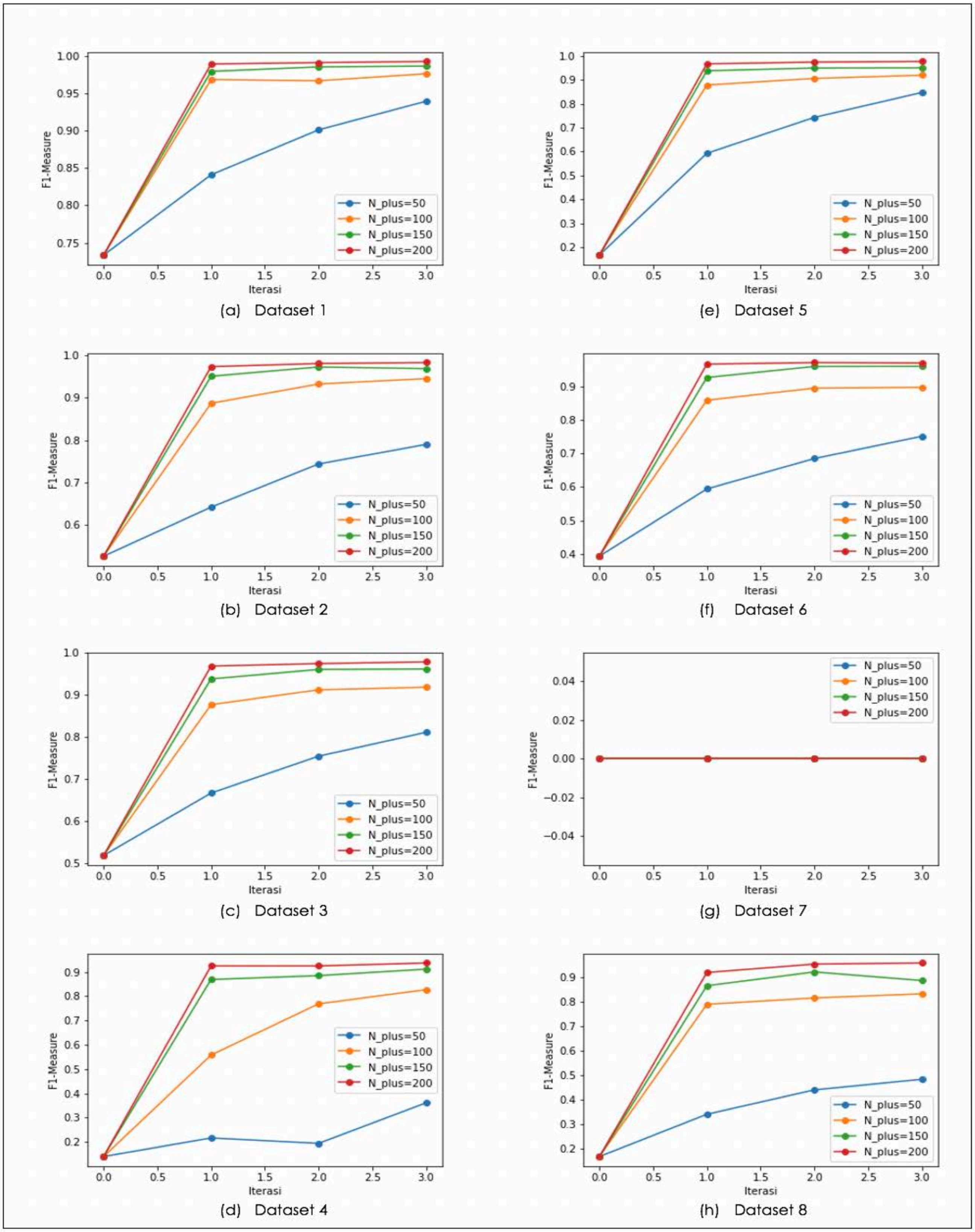
(c) Dataset 3

(g) Dataset 7

(d) Dataset 4

(h) Dataset 8

Fig. 3. Pseudo-Labeling F1-Measure Result..



(a) Dataset 1

(e) Dataset 5

(b) Dataset 2

(f) Dataset 6

(c) Dataset 3

(g) Dataset 7

(d) Dataset 4

(h) Dataset 8

Fig. 4. Performance Comparisons for Cost-sensitive XGBoost.

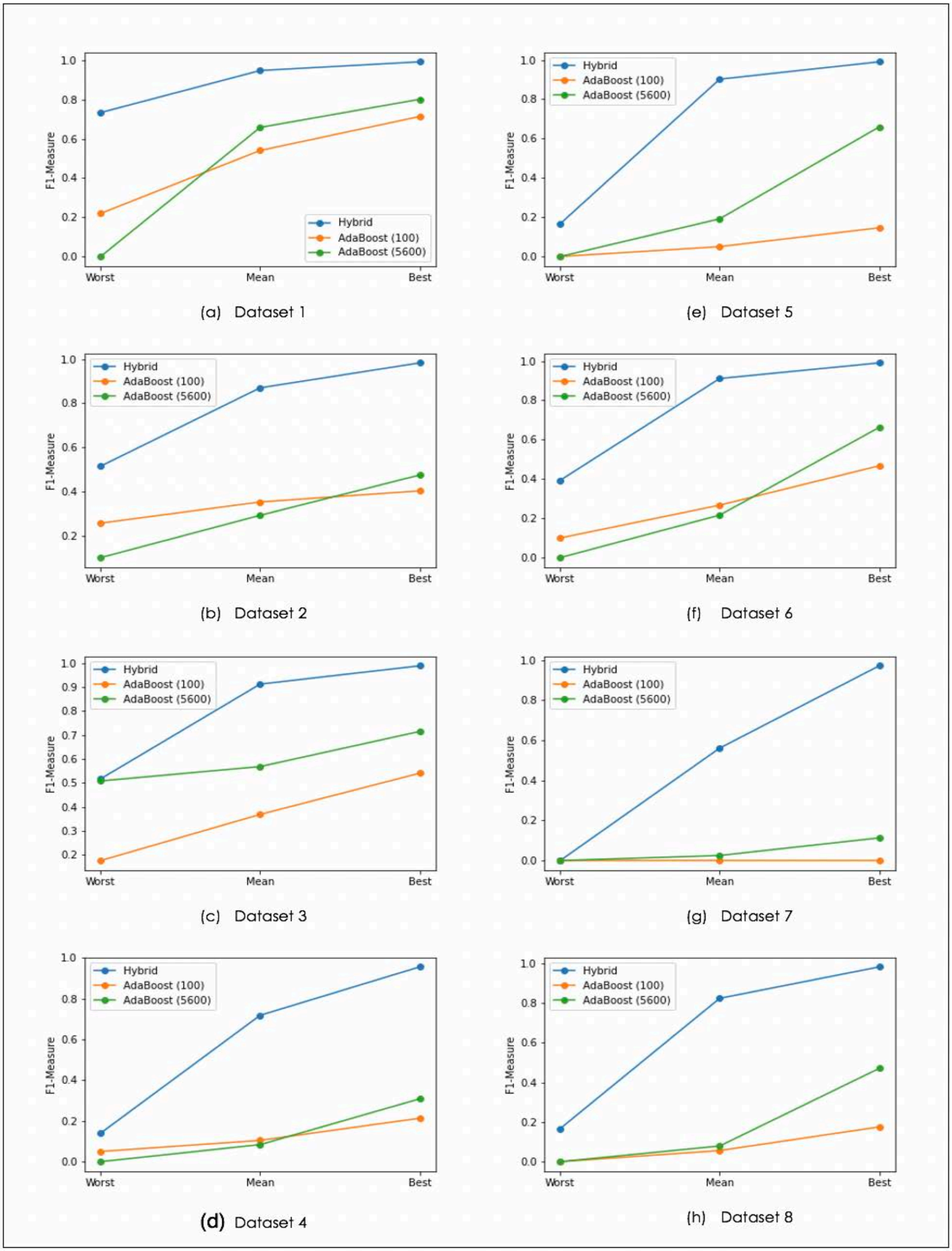


Fig. 5. Performance Comparisons for Hybrid and AdaBoost.

Fig. 5 (a)-(h) are graphs comparing three models, namely the Hybrid Model using 100 labeled data, the AdaBoost Model using 100 labeled data, and the AdaBoost Model using 5600 labeled data. In this comparison, three F1-Measure values are compared for each case, namely the best, mean, and worst F1-Measures that can be obtained from the use of these algorithms.

Based on Fig. 5 (a)-(h), it is found that the best and mean F1-Measure of the Hybrid Model is always higher than the two AdaBoost Models. Overall, the worst F1-Measure of the Hybrid Model is also higher than the two AdaBoost Models. There is only one worst F1-Measure value of the Hybrid Model which is lower than the AdaBoost Model using 5600 labeled data, namely the Dataset 3, but the difference in F1-Measure values is less than 6%. By saving 5500 data labeling, the difference between the F1-Measure values can be neglected. Overall, even though one of the AdaBoost Models already uses 56 times more data, it still cannot compete the F1-Measure of the Hybrid Model.

V. Conclusion

For the 0th iteration, the dataset having a ratio of the amount of "right" labeled data with the amount of "right" labeled data is close to 1. In other words, a balanced dataset or a ratio that is more than 1 produces a model with better performance. Thus, the selection of training data at an early stage must pay attention to this ratio. In addition, the use of the Hybrid Method on these datasets can save labeled data 56 times compared to the AdaBoost Method. The positive class weight parameter has no effect on the performance of the resulting model. The Pseudo-Labeling process with Self Training is able to handle the problem of limited training data, except for the Income Dataset Residents who have F1-Measure with a value of 0% both before and after the Pseudo-Labeling process with Self Training. Hybrid model which is able to produce F1-Measure more than 95.6%, so it can be concluded that the Hybrid Method combines the XGBoost and Pseudo-Labeling Cost-Sensitive Classification with Self Training is able to overcome the problem of unbalanced datasets and data limited label.

Acknowledgment

This work is supported by Pusat Asesmen dan Pembelajaran (PUSMENJAR), Ministry of Cultural, Education, and Research Technology, Republic of Indonesia for providing datasets. The paper is part of the Master Thesis by the 1st author in Magister of Computer Science, Faculty of Mathematic and Natural Sciences, Universitas Gadjah Mada.

References

[1] Valenti, S., F. Neri, and A. Cucchiarelli, An Overview of Current Research on Automated Essay Grading. *Journal of Information Technology Education: Research.* 2: 319-330, 2017.

[2] Herwanto, G. B., Y. Sari, B.N. Prastowo, M. Riasetiawan, I.A. Bustoni, and I. Hidayatulloh, UKARA: A Fast and Simple Automatic Short Answer Scoring System for Bahasa Indonesia. *Proceeding Book of 1st International Conference on Educational Assessment and Policy.* 2: 1-8, 2018.

[3] Riasetiawan, M., B.N. Prastowo, I. Novindasari, and N.J. Aisyiah, Automatic Scoring System for Essay Answer Data using Computational Approach: Clustering and Convolutional Neural. *Prosiding 1st National Conference on Educational Assessment and Policy (NCEAP 2018).* 1: 89-96, 2018.

[4] Fernández, A., S. García, M. Galar, R.C. Prati, B. Krawezyk, and F. Herrera, *Learning from Imbalanced Data Sets.* Switzerland: Springer, 2018.

[5] He, H. and Y. Ma., Imbalanced Learning: Foundations, Algorithms, and Applications. New Jersey: John Wiley & Sons, 2013.

[6] Xia, Y., C. Liu, and N. Liu, Cost-Sensitive Boosted Tree for Lean Evaluation in Peer-to-Peer Lending. *Electronic Commerse Research and Applications,* 2017.

[7] Wang, C., C. Deng, and S. Wang, Imbalance-XGBoost: Leveraging Weighted and Focal Losses for Binary Label-Imbalanced Classification with XGBoost. *Pattern Recognition Letters.* 136: 190-197.

[8] Chen, T., and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Agustus 2016. 785-794, 2020.

[9] Babakhin, Y., A. Sanakoyeu, and H. Kitamura, Semi-Supervised Segmentation of Salt Bodies in Seismic Images using an Ensembles of Convolutional Neural Networks. *Pattern Recognition 41st DAGM German Conference.* Dortmund, Germany. 10-13 September 2019. 218-231, 2019.

[10] Page, E. B, The Imminence of Grading Essays by Computer-25 Years Later. *Computers and Composition.* 10(2): 45-58, 1993.

[11] Xu, R., T. Chen, Y. Xia, Q. Lu, B. Liu, and X. Wang, Word Embedding Composition for Data Imbalances in Sentiment and Emotion Classification. *Cognitive Computation.* 7: 226-240, 2015.

[12] Le, T., M.T. Vo, B. Vo, M.Y. Lee, and S.W. Baik, A Hybrid Approach using Oversampling Technique and Cost-Sensitive Learning for Bankruptcy Prediction. *Complexity.* 2: 1-12, 2019.

[13] Didaci, L., G. Fumera, and F. Roli, Analysis of Co-Training Algorithm with Very Small Training Sets. *Structural, Syntactic, and Statistical Pattern Recognition.* Hiroshima, Japan. 7-9 November 2012. 719-726, 2012.

[14] Peikari, M., S. Salama, S.N. Mozes, and A.L. Martel, A Cluster-then-label Semisupervised Learning Approach for Pathology Image Classification. *Scientific Reports.* 8(1): 7193, 2018.

[15] Pramularsih, M. Cost-Sensitive XGBoost and Psesudo-Labeling with Slef Training for Imbalanced Data and Few Labeled Data in Automated Essay Scoring, Master Thesis in Magister Program of Computer Science, Faculty of Mathematic and Natural Sciences, Universitas Gadjah Mada, Indonesia, 2021.

Prediction of Instructor Performance using Machine and Deep Learning Techniques

Basem S. Abunasser[1], Mohammed Rasheed J. AL-Hiealy[2]
Alaa M. Barhoom[3], Abdelbaset R. Almasri[4], Samy S. Abu-Naser[5]
University Malaysia of Computer Science & Engineering (UNIMY), Cyberjaya, Malaysia[1, 2, 3, 4]
Faculty of Engineering and Information Technology, Al-Azhar University, Gaza, Palestine[5]

***Abstract*—The quality of instructors' performance mainly influences the quality of educational services in higher educational institutions. One of the major challenges of higher educational institutions is the accumulated amount of data and how it can be utilized to boost the academic programs quality. The recent advancements in Artificial Intelligence techniques, including machine and deep learning models, have led to the expansion in practical prediction for various fields. In this paper, a dataset was collected from UCI Repository, University of California, for the prediction of instructor performance. In order to find how effective the instructor in the higher education systems is, a group of machine and deep learning algorithms were applied to predict instructor performance in higher education systems. The best machine-learning algorithm was Extra Trees Regressor with Accuracy (98.78%), Precision (98.78%), Recall (98.78%), F1-score (98.78%); however, the proposed deep learning algorithm achieved Accuracy (98.89%), Precision (98.91%), Recall (98.94%), and F1-score (98.92%).**

Keywords—Education; deep learning; machine learning; prediction; instructor performance

I. Introduction

Machine learning is a subclass of artificial intelligence (AI) [1, 2]. It is concerned with the teaching computers how to learn from different types of data and to enhance with experience without programming explicitly to do that [3]. In machine learning, models are being trained to look for correlations and patterns in huge datasets and be able to make predictions and decisions according to its analysis [4]. Applications of machine learning are improved with usage and become more precise when there are more data at hand. Machine learning applications are everywhere –in our offices, our supermarkets, our social media, and our hospitals [5].

Deep learning are subclass of Machine learning. Deep learning are networks with a great number of layers [6]. These layers can process broad quantities of data and find the weight of associated link in a network; such as, in an image of bird species recognition, part of the layers in the network can detect singular features in the bird's face, such as beak or eyes, whereas another layer could tell if the features in some way designate bird face [7]. Deep learning emulates how human brain operate. A few examples of deep learning are self-driven cars, medical diagnoses through sounds, classification of fish species, detection of different diseases from the eyes of the person [8]. If a network has more layers, these layers can perform complex tasks. Deep learning algorithms need high computer power to be able to produce results [10, 13].

Higher education systems claim new methodologies that increase the achievement, quality and efficiency [14, 15]. Typically Machine and Deep learning algorithms are applied in higher education for examining the effect of educational approaches on students, and in what way students comprehend the course material [11, 22]. The academic performance of students usually is based on some features like the Cumulative Grade Point Average (CGPA), economic situation, demographic data, family background and the models for prediction. Therefore the majority of the research in this field depends on the attributes that are related to the students [12, 21].

This paper is an attempt to analyze the data associated with the evaluation of the student for instructors to enhance the quality of higher educational systems and specify the factors that impact the performance of the students. Student performance prediction is largely associated with the quality of teaching.

In this study, various data prediction techniques are carried out on the student evaluation dataset for the prediction of student accomplishment, inspect instructors' performance, and discover the best technique for classification in line with these measures: Accuracy, Precision, Recall, F1-score and time performance [16].

II. Literature Review

It is required to measure the instructors' performance to boost the effectiveness of teaching and enhance the knowledge of students in the field of higher education. It is done by using feedback gathered from the students.

Agaoglu [1] measured the performances of instructor dependent on view of student using questionnaire of a course evaluation using four techniques: Support Vector Machines, Decision Tree algorithms (C5.0, CART), Discriminant Analysis and Artificial Neural Networks (ANN-2QH, ANN-3QH, ANNMH). The performance measurements applied were recall, precision, accuracy, and specificity. Also, feature importance was done to eliminate the irrelevant features. At last this work indicated the expressiveness and adequacy of models of data mining in higher education. Dataset was collected from Marmara University, Istanbul, Turkey. The dataset contains 2850 records, 25 features, and one class name. Among various strategies C5.0 achieved high accuracy of 92.3%.

The research in Ahmadi and Ahmad [2], inspected the attributes of instructors training performance utilizing two techniques: stepwise regression and decision trees. The dataset in that study was collected from learners of MIS department during the time of 2004-2009. Factor investigation was applied to reveal free factors influencing the overall performance of instructors. Stepwise regression model was created using SPSS and decision trees were built using Answer Tree.

Ahmed et al. [3] inspected the parts which are mainly influencing the success of learners for predicting the performance of instructors to upsurge the quality of the educational system using several techniques like Multilayer Perception (MLP), J48 Decision Tree, Sequential Minimal Optimization (SMO) and Naïve Bayes (NB). The Dataset was collected from (UCI) Repository. There were 32 features collected from Q1 to Q28 asked with responses from 1 to 5. Features were assessed utilizing R, eight features were chosen with high impact. Algorithms were applied with all features and with strongly affected features only. J48 achieved 84.8% with all the features and SMO achieved 85.8% for chosen features.

Ola and Sellapan [16] examined the feedback from students about the instructors to form Instructor Evaluation framework using WEKA tool. Data collected from 830 undergraduate studies around 104 records with five attributes. Decision tree algorithm applied and the outcomes were utilized by the educationalist to distinguish whether specific instructor is proceeded to the following semester or not. The researchers used an intelligent approach for the assessment of instructors' performance in higher institutions and proposed an optimal machine learning algorithm to design a system framework. Formative and Summative assessment methods applied to assess the instructor's performance to increase the quality.

Kumar and Saurabh [17] created a framework to predict the performance of instructors utilizing their assessment, checking the classes and performance assessment of instructors. The strategies utilized in that system was Naive Bayes, ID3, LAD tree and CART in WEKA tool. Three years of data gathered from post graduate students with 14 attributes. The precision created by ID3 was 65.14%, 72.32% by CART, 75.00% by LAD Tree and the most accuracy 80.35% produced by NB classifier.

In the study of Vijayalakshmi et al. [19] some of the machine learning algorithms were applied like "Naïve Bayes, K-Nearest Neighbor, Random Forest, Support Vector Machine, and Decision Tree". The implementation language is R programming language for data mining apps. Various implementation measures were applied to assess the system such as accuracy, precision, recall, specificity, sensitivity. The highest accuracy attained was by SVM. It was better than other models using the dataset at hand.

Yahya et al. [20] examined the practicality of applying Data Mining techniques to distinguish the practicality of instructors. Data was gathered, nine methods were applied such K-Nearest Neighbors (KNN), Naive Bayes (NB), Support Vector Machine (SVM), RA, J48, JPip, AdaBoost, BN, Random Forest among these Random Forest (RF) and SVM showed prominent implementation.

A. *Summary of the Previous Studies*

In Table I, a summarization of the above discussed previous studies in terms of Machine Learning methods, best method, tools used, accuracy of each method used, and the size of the dataset and number of attributes.

Therefore, in the current, the same dataset as in [3] which was collected from UCI Machine Learning Repository from the University of California for the prediction of instructor performance [18] was used.

TABLE I. SUMMARY OF PREVIOUS STUDIES

Reference	**Techniques used**	**Tool Used**	**Results**		**Dataset**
Agaoglu [1]	ANN, DA,C5.0,CART, SVM	IBM SPSS Modeler-Quick and Multiple Classifiers	C5.0	92.3	2850 data, 26 Attributes
			CART	89.9	
			SVM	91.3	
			ANN-Q2H	91.2	
			ANN Q3H	90.8	
			ANN-M	90.5	
			DA	90.5	
Ahmadi and Ahmad [2],	Decision tree with J48	WEKA	J48	82.60	104 records 5 features
Ahmed *et al.* [3]	MLP, DT, NB, ETR	WEKA	DT	84.8	5820 data, 33 Attributes
			NB	83.3	
			ETR	84.5	
			MLP	82.5	
Asanbe *et al.* [8]	MLP, ID3, C4.5	WEKA	ID3	71.00	2010-2015 data, 350 records, 12 Attributes
			C4.5	83.5	
			MLP	82.5	

Ola and Sellapan [16]	Decision tree with J48	WEKA	J48	71.20	830 undergraduate studies around 104 records with 5 attributes
Kumar and Saurabh [17]	NB, ID3, CART, LAD	WEKA	NB	80.35	3 Years data 14 Attributes
			ID3	65.17	
			CART	72.32	
			LAD	75.00	
Vijayalakshmi *et al.* [19]	NB, KNN, RF, SVM, C5.0	R	NB	87.7	2220 data, 21 Attributes
			KNN	91.7	
			C5.0	94.2	
			RF	98.09	
			SVM	99.25	
Yahya *et al.* [20]	KNN , NB, SVM, RA, J48, JPip, AdaBoost, BN, RF	WEKA	KNN	57.90	7348 questions 6 attributes
			NB	57.40	
			SVM	70.80	
			RA	64.90	
			J48	72.60	
			JPip	25.00	
			AdaBoost	64.70	
			BN	20.10	
			RF	55.70	

III. Methodology

This section will present the methodology of our study which includes Data Collection, Data Preprocessing, Explanation of the proposed models used for analysis and prediction.

A. *Dataset*

Dataset used in this study was collected from UCI Repository for the prediction of instructor performance [18]. It has 5820 with 33 features.

The features in the dataset: "instr code, class level, number of repeating the course, attendance, difficulty level, and 28 question (Q1 to Q28). Q1-Q28 are all Likert-type, meaning that the values are taken from 1-5, where 1,2,3,4,5 represents 'Poor', 'Fair', 'Good', 'Very Good', and 'Excellent' respectively for Q1 to Q28. Furthermore, there is one class variable (Performance). Performance was calculated by taking the average of the 28 Question values. The calculated values of these questions are in the form {1, 2, 3, 4, 5}, where 1,2,3,4,5 represents the 'Poor', 'Fair', 'Good' 'Very Good', and 'Excellent' respectively." [18]

B. *Data Analysis*

The class (performance) variable predicts the performance of the instructor. The possible values for performance are 'Poor' (888 Observations),".Fair" (996 Observations), Good" (2073 Observations), "Very Good" (1275 Observations), 'Excellent' (588 Observations). The class (performance) variable distribution is shown in Fig. 1.

C. *Data Preparation*

All of the 33 features and the Performance class of the dataset are already label encoded. The class (Performance) balancing was checked and found that the class is not balanced as in Fig. 1. So, Smote function was used to balance the class (Performance). The Smote function increases the number of samples of the low counts to be the same as the higher count.

D. *Dataset Splitting*

The dataset was split into 3 datasets: (Training, testing and validating datasets). The ratio of splitting was (60%, 20%, and 20%).

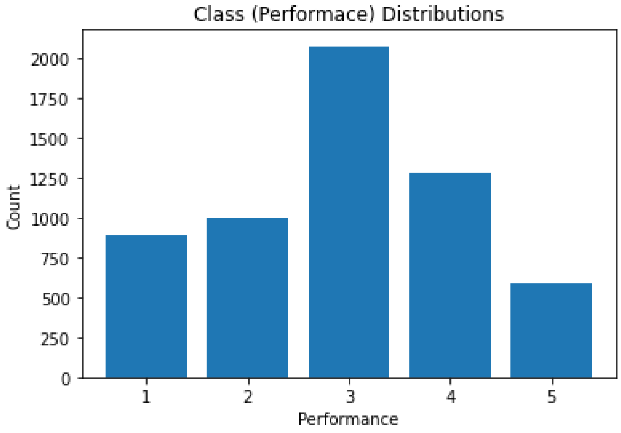


Fig. 1. Class (Performance) Distribution.

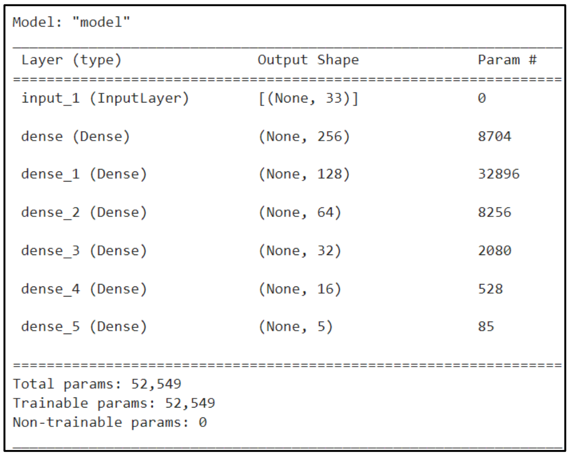
E. *Description of Models used in the Study*

There are many algorithms of ML that can be used in the prediction of instructor Performance level. ML algorithms were trained and tested using the current dataset with 18 various features. The algorithms that were used for prediction and analysis belong to 10 categories of Machine Learning [9] including:

- Naive_Bayes (GaussianNB).
- Neighbors (NearestCentroid, KNeighborsClassifier).
- Linear_model(LogisticRegression,LogisticRegressionCV , LinearRegression).
- SVM (SVC).
- Tree (ExtraTreeClassifier, DecisionTreeClassifier).
- XGBoost (XGBClassifier).
- Ensemble (GradientBoostingClassifier, GradientBoostingRegressor, AdaBoostRegressor, Extra TreesRegressor, BaggingClassifier, RandomForest Classifier).
- Neural_Network(MLPClassifier, MLPRegressor).
- Lightgbm(LGBMClassifier).
- Semi_supervised(LabelPropagation).

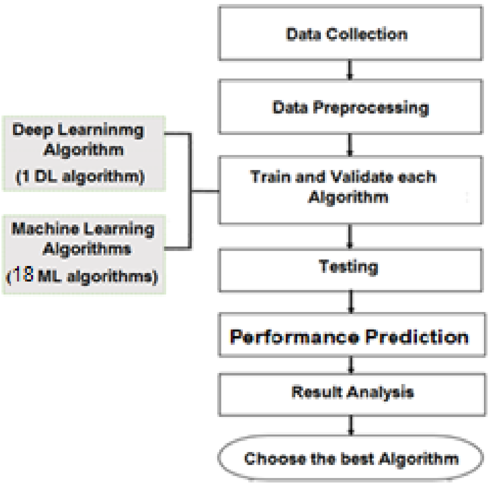
Furthermore, a deep learning model was proposed to predict instructors' performance in higher education systems. The DL proposed model consists of seven Dense layers: one input layer (33 features), five hidden layers (256,128, 64, 32, and 16 neurons), and one output layer with five classes and softmax function as can be seen in Fig. 2. The reason for using five hidden layers is the high accuracy. The structure of the DL model gave the best accuracy compared to four, three, or two hidden layers.

The steps of the methodology used in the study for predicting instructors' performance in higher education systems are summarized in Fig. 3.



```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 33)]              0

 dense (Dense)               (None, 256)               8704

 dense_1 (Dense)             (None, 128)               32896

 dense_2 (Dense)             (None, 64)                8256

 dense_3 (Dense)             (None, 32)                2080

 dense_4 (Dense)             (None, 16)                528

 dense_5 (Dense)             (None, 5)                 85

=================================================================
Total params: 52,549
Trainable params: 52,549
Non-trainable params: 0
_________________________________________________________________
```

Fig. 2. Structure of the Proposed Deep Learning Model.



Data Collection
Data Preprocessing
Deep Learninmg Algorithm (1 DL algorithm)
Machine Learning Algorithms (18 ML algorithms)
Train and Validate each Algorithm
Testing
Performance Prediction
Result Analysis
Choose the best Algorithm

Fig. 3. Methodology for the Prediction of Instructors' Performance in Higher Education Systems.

IV. Result and Analysis

In the following sections, a discussion of the result achieved by the deep learning model and the machine learning models will be presented. The first section talks in detail about the Performance Evaluation and the second section presents the Performance Analysis of all models used in this study.

V. Performance Evaluation

Different machine and deep learning measurements can be applied on the various model used in the current study. The most popular measurements are: Accuracy, F1- score, Recall and Precision are the most important criterion used to assess a models performance. The value of the confusion matrix which is generated during the testing of the model is considered to calculate these measurements as illustrated in equation 1, 2, 3 and 4.

$$\text{Accuracy} = (TP + TN) / (TP+TN+FP+FN) \quad (1)$$

$$\text{Precision} = (TP / (TP+FP) \quad (2)$$

$$\text{Recall} = (TP / (TP+FN) \quad (3)$$

$$F1 = 2 \text{ x } (\text{Precision x Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

Where TP = True Positive, TN = True Negative

FP = False Positive, FN = False Negative

VI. Performance Analysis of Applied Models

In this study, 18 machine learning algorithms to predict instructors' performance in higher education systems were used. Furthermore, a deep learning model was proposed to predict instructors' performance in higher education systems. The aim of this study was to get a more efficient predictive model by making a comparison between the different deep and machine learning models. 60% of the dataset for training, 20% of the dataset for validating and the remaining 20% of the dataset were used for the testing.

TABLE II. PERFORMANCE OF THE MACHINE AND DEEP LEARNING ALGORITHMS

Model Type	Model Name	Accuracy	Precision	Recall	F1-score	Time in Sec
Machine Learning	Extra Trees Regressor	**98.78%**	**98.78%**	**98.78%**	**98.78%**	2.70
	Gradient Boosting Regressor	97.71%	97.71%	97.71%	97.71%	0.80
	Random Forest Classifier	97.35%	97.36%	97.35%	97.35%	0.65
	Logistic Regression CV	97.25%	97.25%	97.25%	97.25%	308.06
	LGBM Classifier	97.01%	97.01%	97.01%	97.01%	1.10
	Gradient Boosting Classifier	95.95%	95.95%	95.95%	95.95%	5.55
	MLP Classifier	95.80%	95.80%	95.80%	95.80%	11.37
	Bagging Classifier	95.75%	95.75%	95.75%	95.75%	0.25
	Logistic Regression	94.64%	94.64%	94.64%	94.64%	1.93
	Extra Tree Classifier	93.83%	93.83%	93.83%	93.83%	0.02
	Decision Tree Classifier	93.39%	93.39%	93.39%	93.39%	0.03
	Label Propagation	93.25%	93.25%	93.25%	93.25%	2.02
	Gaussian NB	92.47%	92.47%	92.47%	92.47%	0.02
	K Neighbors Classifier	91.66%	91.66%	91.66%	91.66%	0.41
	SVC	91.66%	91.66%	91.66%	91.66%	0.61
	Linear Discriminant Analysis	91.27%	91.27%	91.27%	91.27%	0.07
	Ada Boost Regressor	91.16%	91.16%	91.16%	91.16%	2.12
	Nearest Centroid	90.79%	90.79%	90.79%	90.79%	0.02
Deep Learning	Proposed Deep Learning Model	**98.89%**	**98.91%**	**98.94%**	**98.92%**	**2.00**

To evaluate the models performance, five sorts of assessment measures were engaged: "Recall, Precision, Accuracy, F1-Score and time needed for each model to run are shown in Table II. It is observed that the best machine-learning algorithm was "Extra Trees Regressor" with an Accuracy (98.78%), Precision (98.78%), Recall (98.78%), and F1-score (98.78%); however, the proposed deep learning algorithm achieved an Accuracy (98.89%), Precision (98.91%), Recall (98.94%), and F1-score (98.92%).

VII. RESULT AND DISCUSSION

All previous studies reviewed in the section of literature review except one used different datasets; thus the results of these studies cannot be compared with the result obtained in the current study.

The previous study that used the same dataset as in the current study is Ahmed et al. [3]. The following table compares their results with the current proposed model's results.

As it can be seen in Table III, the results of the current study are much higher than the results obtained in the previous study.

TABLE III. RESULTS COMPARISON WITH PREVIOUS STUDIES

Model Name	**Previous study (Ahmed *et al.* [3])**	**Proposed Models of the Current Study**
Decision Tree Classifier (DT)	84.80	93.83
Gaussian NB	83.30	92.47
Extra Trees Regressor (ETR)	84.50	98.78
MLP Classifier	82.50	95.80

The reason for the high accuracy of the current study is the pre-processing the handling of the dataset.

VIII. CONCLUSION

In this study, 18 different Machine Learning algorithms and one deep learning algorithm for predicting instructors' performance in higher education systems were used. The dataset was collected from UCI Repository for the prediction of instructor performance. The dataset was preprocessed, the class (performance) was balanced using smote function. Each algorithm was trained, tested and its performance was noted. Furthermore, the proposed deep learning model was trained, validated and tested using the same dataset and its performance was noted. Among all the machine learning models used, the best machine-learning algorithm was Extra Trees Regressor with an Accuracy (98.78%), Precision (98.78%), Recall (98.78%), F1-score (98.78%); however, the proposed deep learning algorithm achieved an Accuracy (98.89%), Precision (98.91%), Recall (98.94%), F1-score (98.92%). Even though, the accuracies of the best machine learning algorithm and the proposed deep learning algorithm were close; the proposed deep learning algorithm was slightly better.

These discoveries are helpful to educationalist to improve their performances.

REFERENCES

[1] Agaoglu, Mustafa. "Predicting Instructor Performance Using Data Mining Techniques in Higher Education." IEEE Access, 2016, Vol. 4, pp. 2379-2387.

[2] Ahmadi, Fateh, and ME Shiri Ahmad. "Data Mining in Teacher Evaluation System using WEKA." International Journal of Computer Applications, 2013, Vol. 63, No. 10.

[3] Ahmed, Ahmed Mohamed, Ahmet Rizaner, and Ali Hakan Ulusoy. "Using data mining to predict instructor performance." Procedia Computer Science, 2016, Vol. 102, PP. 137-142.

[4] Saleh, A., Sukaik, R., Abu-Naser, S.S. Brain tumor classification using deep learning. Proceedings - 2020 International Conference on Assistive and Rehabilitation Technologies, iCareTech 2020, 2020, pp. 131–136, 9328072.

[5] Arqawi, S., Atieh, K.A.F.T., Shobaki, M.J.A.L., Abu-Naser, S.S., Abu Abdulla, A.A.M. Integration of the dimensions of computerized health information systems and their role in improving administrative performance in Al-Shifa medical complex, Journal of Theoretical and Applied Information Technologythis link is disabled, 2020, Vol. 98, No. 6, pp. 1087–1119.

[6] Buhisi, N. I., & Abu-Naser, S. S. Dynamic programming as a tool of decision supporting. Journal of Applied Sciences Research, 2009, Vol. 5, No. 6, pp. 671-676.

[7] Naser, S. S. A. Developing visualization tool for teaching AI searching algorithms. Information Technology Journal, 2008, Vol. 7, No. 2, pp. 350-355.

[8] AlKayyali, Z. K. D., et al. Prediction of Student Adaptability Level in e-Learning using Machine and Deep Learning Techniques. International Journal of Academic and Applied Research (IJAAR), 2020, Vol. 6, No. 5, pp.84-96.

[9] Naser, S. S. A. Intelligent tutoring system for teaching database to sophomore students in Gaza and its effect on their performance. Information Technology Journal, 2006, Vol. 5, No. 5, pp. 916-922.

[10] Mady, S.A., Arqawi, S.M., Al Shobaki, M.J., Abu-Naser, S.S. Lean manufacturing dimensions and its relationship in promoting the improvement of production processes in industrial companies. International Journal on Emerging Technologies, 2020, Vol. 11, No. 3, pp. 881–896.

[11] Albatish, I.M., Abu-Naser, S.S. Modeling and controlling smart traffic light system using a rule based system. Proceedings - 2019 International Conference on Promising Electronic Technologies, ICPET 2019, 2019, pp. 55–60.

[12] Elzamly, A., Messabia, N., Doheir, M., ...Al-Aqqad, M., Alazzam, M. Assessment risks for managing software planning processes in information technology systems. International Journal of Advanced Science and Technology, 2019, Vol. 28, No. 1, pp. 327–338.

[13] Abu Ghosh, M.M., Atallah, R.R., Abu Naser, S.S. Secure mobile cloud computing for sensitive data: Teacher services for palestinian higher education institutions. International Journal of Grid and Distributed Computing, 2016, Vol. 9, No. 2, pp. 17–22.

[14] Elzamly, A., Hussin, B., Naser, S.A., ...Selamat, A., Rashed, A. A new conceptual framework modelling for cloud computing risk management in banking organizations. International Journal of Grid and Distributed Computing, 2016, Vol. 9, No. 9, pp. 137–154.

[15] Abu-Naser, S.S., El-Hissi H., Abu-Rass, M., & El-khozondar, N. An expert system for endocrine diagnosis and treatments using JESS. Journal of Artificial Intelligence, 2010, Vol. 3, No. 4, pp. 239-251.

[16] Abu Naser, S.S. Evaluating the effectiveness of the CPP-Tutor, an intelligent tutoring system for students learning to program in C++. Journal of Applied Sciences Research, 2009, Vol. 5, No. 1, pp. 109-114.

[17] V. Vijayalakshmi, K. Panimalar, S. Janarthanan. “Predicting the performance of instructors using Machine learning algorithms.” High Technology Letters, 2022, Vol. 26, No. 12, pp. 694-705.

[18] UCI Machine Learning Repository, https://archive.ics.uci.edu/.

[19] Zaqout, I., et al. "Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology." International Journal of Hybrid Information Technology, 2015, Vol. 8, No. 2, pp. 221-228.

[20] Obaid, T., Eneizan, B., Naser, S.S.A., ...Abualrejal, H.M.E., Gazem, N.A. Factors Contributing to an Effective E- Government Adoption in Palestine. Lecture Notes on Data Engineering and Communications Technologies, 2022, 127, pp. 663–676.

[21] Naser, S. S. A. JEE-Tutor: An intelligent tutoring system for java expressions evaluation. Information Technology Journal, 2008, Vol. 7, No. 3, 528-532.

[22] Naser, S. S. A. Developing an intelligent tutoring system for students learning to program in C++. Information Technology Journal, 2008, Vol. 7, No. 7, pp. 1051-1060.

Application-based Usability Evaluation Metrics

Hanaa Bayomi[1], Noura A.Sayed[2], Hesham Hassan[3], Khaled Wassif[4]
Faculty of Computers and Artificial Intelligence, Cairo University, Cairo, Egypt[1, 3, 4]
Faculty of Computers and Artificial Intelligence, Akhbar El Youm Academy, Cairo, Egypt[2]

***Abstract*—Testing is one of the vital stages in the software development life cycle (SDLC). Usability testing is a very important field that helps the applications be easily used by the end-users. Because of the importance of usability testing, a metrics has been developed to help in measuring the usability through converting the main qualitative usability attributes in ISO to quantitative steps that provide the developer a framework to follow in developing to achieve usability of their applications and helps the tester with a checklist and a tool to measure the usability percentage of their application. The framework provides a set of steps to achieve the usability attributes and answers the question of how you could measure this attribute with the defined steps. The framework results in a 95% average accuracy in the high-rate application and a 59% average accuracy in the low-rate application. Finally, the framework is programmed in a tool to measure the usability percentage of the application through a checklist and provides a scheme to help the developer achieve the best results in usability.**

***Keywords*—Usability; human-computer interaction; evaluation; quantitative attributes; testing**

I. Introduction

Testing is one of the important stages of the SDLC. Usability testing is one of the non-functional testing types [1]. Usability is one of the quality dimension that evaluates the quality and usefulness of applications [2]. Usability is a quality attribute that determines how easily and simply user interfaces can be used. So, usability is the interface’s necessary condition for application’s survival [3]. If an interface is difficult to use and the users get lost, they will leave. No user would be bothered with reading a manual first before using an interface or a website. Therefore, usability is a very vital element of an application interface.

Accordingly, a framework has been developed to convert the qualitative usability attributes into qualitative steps to help the developer to have simple steps to achieve the highest usability and helps the tester to have a checklist to easily evaluate the application usability. According to Nielson, usability can be defined as a method for improving the design process [4]. Usability is assessed based on six dimensions, which include learnability, memorability, efficiency, effectiveness, error rate, and user satisfaction. [5]. Usability testing uses the black box testing technique [1]. Our framework can be performed by the tester at the system and acceptance levels. As well, usability is one of the nonfunctional requirements of any software, which is one of the core areas of research in the field of human-computer interaction [6].

Now-a-days, both Web and mobile applications are considered the most two popular types of applications. Usability has been defined as a key component in the overall quality of a software product, and research shows that usability can determine the success or failure of a software system [7]. Usable software systems are not only more efficient, accurate, and safe, but also much more successful and several studies have shown the benefits of incorporating usability evaluation in the process of software development. Therefore, usability evaluation has become an important research field [7].

The remainder of this paper is organized as follows: Section II is related work, which provides background information about usability and testing techniques; Section III is methods, which describe how the research is designed and carried out; Section IV is usability attributes, which describe the attributes in well-defined steps; Section V is testing references and questions, which describe the usability attributes in testing questions; and Section VI is results, which describe the findings.

II. Related Work

Nielsen (1994) defined usability as a quality attribute that measures how user interaction can be used as a method for improving ease-of-use, efficiency, and satisfaction [4]. Standard ISO 9241-11 defines usability as "the extent to which a product can be used by specific users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [7]. It was further elaborated that there are other important concepts related to usability, such as user, goal, effectiveness, efficiency, satisfaction, and context of use [8].

Usability of software applications is one of the core areas of research in the field of human-computer interaction. There were many methods to inspect, test, and inquire about an application's usability [7], [9].

Making the application usable doesn’t mean that it will be easy. Some applications must be a little bit difficult to use to help in reaching the target of the application's usage for example e-learning applications [10]. The e-learning applications' difficulty must be specified according to the learning level of the student to meet the target of increasing the student's learning curve [11]. One of the very effective ways of evaluating the e-learning applications’ usability is the heuristic evaluation method [12], [13]. Heuristic evaluation can also be used to measure other types of applications because it has been shown to be effective in inspecting errors [14], [15], [16].

Before evaluating any application, the latest issues must be reviewed and the most commonly occurring errors arise related to your application type [17]. To change the qualitative attributes to quantitative, the quantitative attributes must be revised [18]. Mobile and web applications are the most commonly used types of applications today. Both of those types can be accessed via mobile, which is the most convenient device to hold most of the time, so we must concentrate on mobile application usability [19], [20], [21], [22], [23].

Most of the papers focus on updating the usability attributes by adding some new attributes or deleting some of existing attributes or changing the way of the attribute evaluation [24]. Some of the papers are updating the usability main categories introduced in the ISO/IEC 25022 [3], [6]. Some papers are creating new modules to be remotely evaluating the applications through user usage [25]. Some papers focus on one of the usability evaluation techniques [26] [27]. Some papers focus on analyzing the Human computer interaction or usability attributes according to a specific application type [28], [29]. There is some approaches and frameworks that improve issues tracking [30], [31], [32].

According to the ISO, those are the documented standards presented in Table I [6].

TABLE I. ISO USABILITY ATTRIBUTES REFERENCES

ISO reference number	Usability attribute category
ISO 9241-11	Usability Definitions and concepts, replacing the 1988 version of ISO 9241-11
ISO 9241-220	Processes for enabling, executing and assessing human-centered design within organizations, replacing the earlier ISO TR 18529
ISO/IEC 25066	Common industry Format for Usability — Evaluation Reports
ISO/IEC 25022	Measurement of quality in use, (includes measures of effectiveness, efficiency and satisfaction), replacing ISO TR 9126-4
ISO/IEC 25023	Measurement of system and software product quality, (includes measures for usability attributes), replacing ISO/IEC TR 9126-2 and ISO/IEC TR 9126-3

This research highlighted the following usability gaps: first, there is a problem in evaluating the usability of the application's interface. Second, developers do not focus on the simplicity of the interface as they focus on the functionality of the application because the functionality is easy to develop as there are no clear guidelines for evaluating the interface usability. Third, there is no balance between user experience and interface simplicity. For example, learning applications mustn't be so easy or so difficult to meet the goal of increasing the learning curve. Finally, there is no clear qualitative borderline to be applied by the developer and tested by the tester. To fill those gaps, a framework has been developed to describe the usability attributes in fine steps and helps in achieving the best usability of the application.

III. METHODOLOGY

To convert usability attributes in the ISO from qualitative to quantitative steps, this is done by searching for how the developer could achieve each attribute in well-defined steps and asking usability experts what the points they take into consideration to make the application usable enough for the user. The 20 developers and testers were asked about the frequently repeated problems that they face while developing and testing the applications. After collecting the required data, usability attributes have been chosen from the ISO Usability standards. Some of those attributes were merged and some have been added to create an effective usability framework. In the framework, the qualitative chosen attributes have been converted to detailed steps to help the developer and the tester in their practical work. Then, one or two questions is assigned for every attribute to be asked by the tester for evaluation and tested on the application to measure the usability.

The results have been developed by choosing two applications, one with a high rate (amazom) and the other with a low rate (waradly) to test the effectiveness of the framework developed. The first is Amazon, which has a global web application and also a mobile application with a rate of 4.3 out of 5, and the other is Waradly, which is an Egyptian e-commerce site with an intermediate rate, but they don't have a mobile application. To test our framework, three main tasks have been chosen that are performed by any e-commerce application, which are searching, seeking product data, and purchasing.

Finally, a tool has been developed using our framework to measure the usability percentage of the application through a checklist and provide a scheme to help the developer to achieve the best results in usability (http://abuem.com/).

IV. USABILITY ATTRIBUTES

Accordingly, steps have been assigned for each attribute in each category to be measured quantitatively, merging, and adding some attributes to them.

A. *Effectiveness*

1) Tasks completed and objectives achieved: The applications mainly provide the user with services or products. To complete the task, you have to use a scale from 1 to 10 according to the feasibility analysis conducted by the business analyst. To achieve effectiveness in tasks completed and objectives achieved, set reasonable objectives according to each business environment, choose the best developing environment to achieve these objectives (desktop, mobile, or web application), set tasks, specify for each task input and output, specify for each task severity and priority, and order the tasks according to the scale set by feasibility analysis. Here is an example to calculate the rest %. Subtract the finished tasks from the number of all tasks to give you the rest of the tasks. Then divide the result by the number of all tasks and multiply by 100 to give you the whole percent of the rest of the tasks.

2) Errors in a task: To achieve effectiveness in errors in a task attribute, the errors that may arise at each task individually must be predicted; choose an efficient programming language to minimize the predicted errors, fix the unpredicted errors, and document them with their solution to be fixed automatically if they arise again, according to the literature.

3) Tasks with errors: To achieve the effectiveness in tasks with errors attribute, make sure that the errors in each stage will be solved individually, don't move to the next task unless the previous task is clear of any errors, and after testing, the total number of errors mustn't exceed 25% of all the tasks to deliver the system to the end-user.

4) Task error intensity: To achieve the effectiveness in task error intensity attribute, task errors must be classified according to task complexity, error intensity must be set from 1 to 5 based on the severity and priority set by a tester, and the factors that cause the error must be identified in order to solve it or give the user a hint on how to deal with it in order to reduce the intensity of its existence.

B. Efficiency

1) Task time: To achieve the Efficiency in Tasks time attribute, the developer must set the shortest way for the user to make the task to achieve the goal in the easiest way**, s**et the minimum developing time for a task according to its complexity without the existence of any errors**, s**et the maximum developing time for a task according to its complexity with the existence of all errors that could happen, and complete each task within the maximum and minimum of it.

2) Time efficiency: To achieve the Efficiency in Tasks efficiency attribute, the developer must choose a task at a time to set its input and desired output. Set the best programming language to achieve your target and choose the best algorithm according to its big O to achieve the best time efficiency.

3) Cost-effectiveness: To achieve the efficiency in cost-effectiveness attribute, a feasibility study must be developed based on the hardware and software needs (developing and testing environment) and choose the best platform to satisfy user needs, including user supplies to support most of the devices he can afford in software and hardware.

4) Productive time ratio: To achieve efficiency in the productive time ratio attribute, the developer must develop a timing strategy for ending the project and create criteria to set time for each feature based on the employee performance to achieve the optimal standard above and more quantity.

5) Unnecessary actions: To achieve the efficiency in unnecessary actions attribute, after finalising your code, all of the code must be revised to make sure that you choose the best development path and choose the quickest way to make the function with the minimum number of steps.

6) Fatigue: To achieve the Efficiency in Fatigue attribute, the developer must be aware of the problems that occur while developing that affect the deadlines and add them to the delivery date, and be prepared for the cases that make the system load slowly or go down.

C. Satisfaction

1) Overall satisfaction: To achieve satisfaction in the overall satisfaction attribute, the project must be delivered before the deadline, have well-defined requirements that have one meaning, and involve small prototypes to involve the users while developing.

2) Satisfaction with features: To achieve the satisfaction with features attribute, continuous messages must be displayed to inform the user of the current processing, block all the expected errors, and minimize the needed resources for working.

3) Discretionary usage: To achieve the Satisfaction in Discretionary usage attribute, the developer must think about all the cases where the user needs to use the application, and according to each case, the developer must satisfy the user's needs, satisfy the client's needs, and satisfy the market's needs. Finally, the developer must develop all the ways to achieve each target, for example, adding a speech-to-text feature and chatting to enter data.

4) Feature utilization: To achieve the satisfaction in feature utilization attribute, each feature must be explained to the user, choose the simplest way to develop each feature; they must choose the shortest number of steps to end a task; and the page must contain at most one objective or task.

5) Proportion of users complaining: To achieve satisfaction in the overall satisfaction attribute, the application must be too clear to reduce user inquires, provide a chat pot as a supporter in any problem, and provide quick support in a large problem.

6) Proportion of user complaints about a particular feature: To achieve the satisfaction in the overall satisfaction attribute, the application must help the user at the beginning by giving him a demo for each feature, giving the feature a definition and descriptive logo, and making the feature complete on one page with fine input and several steps. Then put the main features on the home page and their shortcuts on other pages. Finally, try to use AI in the application to tell the user if he is making the wrong decision.

7) User trust: To achieve the satisfaction in the overall satisfaction attribute, the application must: get permission from the user before saving or using his data, give a full explanation for why their data is needed as audio or text, solve all the user's problems most simply; then support the user with all the knowledge he may need; and involve the user as much as possible while developing to see the prototypes.

8) User pleasure: To achieve the satisfaction in user pleasure attribute, the application must develop every single task easily with fewer errors, give users continuous explanations and motivation messages, choose cheerful designs and colors to make the user happy, and ask the user for their feedback continuously.

9) Physical comfort: To achieve the satisfaction in physical comfort attribute, a backup must be taken for all your work to make the user feel safe, use UPS to assure that the data will be safe, and use the simplest way to develop the application.

D. Appropriateness Recognizability

1) Readability: To achieve the appropriate level of recognizability in the readability attribute, the icons must be readable to allow the other attributes like learnability, accessibility, and so on to be achieved. So, the font of the icon must be as big as you can to minimize the number of objects that exist on one page, whether it is a mobile or a web application.

2) Description completeness: To achieve appropriateness recognizability in the description completeness attribute, user stories must be wisely described to satisfy the developer and have one meaning based on the user's point of view, market needs, and recommendation of the best way to develop (interface architecture).

3) Demonstration coverage: To achieve the appropriateness recognisability in demonstration coverage attribute, the developer should develop many application types and many targets of applications at the same field to be aware of all the vulnerabilities of developing, and while developing, the developer must study all the pros and cons of any way related to developing.

4) Entry point self-descriptiveness: To achieve the appropriateness recognizability in the entry point self-descriptiveness attribute, requirements must be determined based on user stories that describe each scenario that could happen by the user to make the most effective decision while developing, and each feature must be described in the simplest way to make it easy to develop by a beginner developer.

E. Learnability

1) User guidance completeness: To achieve the learnability in user guidance completeness attribute, the project mustn’t be launched until the user manual is available in the form of steps.

2) Entry fields defaults: To achieve the learnability in entry fields defaults attribute, the right option must be very clear to the user to avoid errors in data entry fields.

3) Error message: To achieve the learnability in an error message attribute, every problem in every step the user can do, an error message must be designed to clarify the error and help him avoid it.

4) Understandability: To achieve the learnability in the understandability attribute, every step must be clear and mustn’t have a double meaning as possible.

5) Self-explanatory user interface: To achieve the Learnability in Self-explanatory User Interface Attribute, interface parts must be mapped to the objects used in our lives to be easy to understand, recognize, and recall.

F. Operability

1) Operational consistency: To achieve operability in the operational consistency attribute, the page mustn’t contain many tasks or any other information that doesn’t relate to the main task (each page for one purpose).

2) Message clarity: To achieve the operability in the message clarity attribute, any message that appears to the user must be very clear, and if the developer has time, he must add an option to let the user see the last message again when it disappears.

3) Functional customizability: To achieve the operability in the functional customizability attribute, the application must have fewer mandatory fields in each task as much as possible for less time consumption and make a detailed manual consisting of steps to make the application very easy.

4) User interface customizability: To achieve the operability in the user interface customizability attribute, flexibility must be given to the user to change the places of the objects on the interface as well as the colors to make the application more comfortable.

5) Monitoring capability: To achieve the operability in the monitoring capability attribute, the application must provide the user with an indication when he gets closer to the end of his task and provide the user with full statistics of the application usage and resource consumption. Then, inform the user about memory and resource usage, help the user to reduce memory consumption, and don't ask the user for too much data to be saved or user in one task.

6) Undo capability: To achieve the operability in the undo capability attribute, the user must have the ability to go back with clear action and help them to recognize it in many situations.

7) Understandable categorization of information: To achieve operability in understandable categorization of information attributes, the application must provide the user with a thorough explanation of the input data and why it is required, and then inform the user if his data will be used.

8) Appearance consistency: To achieve operability in the appearance consistency attribute, the system must have one design on every page, stick to simple and realistic symbols and designs to relate them to our lives, avoid attaching unrelated things to the application, and avoid putting too many details on one page even if they are related to the same task.

9) Input device support: To achieve the operability in the input device support attribute, the application must work on cheap resources, don’t make the application need many hardware devices to work properly, and try the application before delivery to the customer on different category devices according to the budget.

G. User Error Protection

1) Avoidance of user operation error: To achieve user error protection in the avoidance of user operation error attributes, the application mustn’t allow the user to input wrong data and must give a full explanation for the user's errors.

2) User entry error correction: To achieve user error protection in the User entry error correction attribute, the application has to give the user a full description of the error type and reason for its happening and give the user suggestions to correct his error.

3) User error recoverability: To achieve user error protection in the user error recoverability attribute, the application must allow the user to go back one step to make the user feel safe, give the user a description for each step he makes to increase awareness, and provide the user with a short manual that explains all the steps of the application.

H. User Interface Aesthetics

To achieve user interface aesthetics in the appearance aesthetics of user interfaces attribute, the developer has to choose the best comfortable color for the eye, the uploaded pictures must be of high resolution and minimum size to decrease the time of loading and increase the font of the data as much as possible to satisfy the user.

I. Accessibility

1) Accessibility for users with a disability: To achieve user interface aesthetics in the appearance aesthetics of user interfaces, the developer must be aware of all the types of users that will use the application and their disabilities, like color blindness, wheelchairs, etc.

2) Supported languages adequacy: To achieve user interface aesthetics in **s**upported languages adequacy attribute, the application must support the global language and the regional language, and in the future, you can add the most important languages ordered by priority.

V. TESTING REFERENCES AND QUESTIONS

Some will need to be revised from the requirements documentation, while others will need to be tested on the application. Here is a table that identifies which attributes need to be tested from the application only and which attributes need to be revised from what has been approved in the requirements documentation and sets some suggested questions for each attribute described in Table II.

TABLE II. DESCRIPTIVE QUESTION OF USABILITY ATTRIBUTES

Attributes	**Requirements documentation or Application usage**	**Questions to ask for an evaluation**
Tasks completed and Objectives achieved	Requirements documentation	Are all the system tasks according to the system requirements is done? Do you achieve all the objectives of the system according to the required documentation?
Errors in a task	Application usage	Do you have any errors in any tasks? Then try it
Tasks with errors	Application usage	After testing each task, is all the tasks that have errors exceed 25% of all the tasks?
Task error intensity	Application usage	Is the task error frequently happening?
Task time	Application usage	Is each task takes the minimum time it must take?
Time efficiency	Application usage	Do you develop each task to be done at the minimum processing time?
Cost-effectiveness	Requirements documentation	Are you stick to the feasibility analysis you made before starting?
Productive time ratio	Requirements documentation	Is every second in the application is useful to the user?
Attributes	Requirements documentation or Application usage	Questions to ask for an evaluation
Unnecessary actions	Application usage	Is every step in each task a must and has a specific target?
Fatigue	Application usage	Taking into consideration the maximum number of users at a time works with stability?
Overall satisfaction	Application usage	Dose all the usability requirements tested to achieve high satisfaction?
Satisfaction with features	Application usage	Does the simplicity of each feature exist?
Discretionary usage	Application usage	Can you use the system anytime anywhere?
Feature utilization	Application usage	Do all the features the system support are mandatory?
The proportion of users complaining	Application usage	Will the user need to call customer support or see the manual many times while usage?
The proportion of user complaints about a particular feature	Application usage	Does every feature need to be seen first from a manual?
User trust	Application usage	Does the system makes the user feel that he/ she is secure enough to trust the system?
User pleasure	Application usage	Does the system make the user pleased by the system?
Physical comfort	Application usage	Does the system allow comfortable positions while usage?
Readability	Application usage	Do all the pages is seen clearly by higher ages?
Description completeness	Application usage	Do you make a user manual to explain the system in steps or develop a demo to help the user to see the steps of any task?
Demonstration coverage	Application usage	Does the system has an explanation coverage?
Entry point self-descriptiveness	Application usage	Is every step described in its meaning, target, and how to the user without using the manual?

User guidance completeness	Application usage	Do you create a manual to help the user in explaining the system?
Entry fields defaults	Application usage	Do you explain the default inputs to the user?
Error message	Application usage	Do you give messages every time the user could make an error?
understandability	Application usage	Is the system easy to use and helps user memory to recover quickly?
Self-explanatory user interface	Application usage	Is the interface explain itself by mapping the tasks to the user knowledge?
Operational consistency	Application usage	Does the page contain many tasks?
Message clarity	Application usage	Does the information in all the messages that appears to the user clearly described?
Functional customizability	Application usage	Do all the fields in every step are mandatory?
User interface customizability	Application usage	Does the user have the flexibility to change the interface appearance?
Attributes	Requirements documentation or Application usage	Questions to ask for an evaluation
Monitoring capability	Application usage	Do the system support monitoring for the user errors to help him in the long run?
Undo capability	Application usage	Do you allow the user to go back whenever he wants?
Understandable categorization of information	Application usage	Do you inform the user of the structure of the data and why the application uses it?
Appearance consistency	Application usage	Do the system parts have consistency?
nput device support	Application usage	oes all the input devices available to the user while working with the system?
voidance of user operation error	Application usage	o you take into consideration all the errors that can be done by the user and block them?
User entry error correction	Application usage	Do you design error messages to help the user with their errors?
User error recoverability	Application usage	Do you allow the user to solve the errors he/she made?
Appearance aesthetics of user interfaces	Application usage	Does the interface explain every step?
Accessibility for users with a disability	Application usage	Is the system helps users with different disabilities to use it easily?
Supported languages adequacy	Application usage	Do the system support many languages mainly the global language and the user's first language?

VI. Results and Discussion

A. Results

Three main tasks have been chosen according to their importance to any e-commerce users which are search, seek product data and purchase to test them on two applications one with high rate (amazom) and the other with low rate (waradly). Assume that the attributes that need to be revised from the requirement documentation are all true. The results have been calculated by testing framework attributes on the two application simultaneously through asking the previous questions of each attribute. The answers will be true or false. If answer is true, so the attribute is applied on the application. If answer is false, so the attribute is not applied on the application. Every true will be converted to 1 grade and every false will be converted to 0 grade. Finally, all the grades of all the attributes are added, then divided by the total number of attributes which are 44, and multiplied by 100 to get the usability percentage of each task usability. For example, after testing the searching task on amazon application by opening the application and answer all the usability attributes on the searching task to check the existence of every attribute, the result was 42 attributes were true from total 44 attributes. The result was calculated by calculating total true attributes, divided by total number and multiply 100 (42/44*100) to give the result 95%. All the final results are presented in Table III after testing all the usability attribute on every e-commerce task then compare the usability results of every application to proof that the framework can differentiate between the usability of the high rate and the low rate applications.

TABLE III. Usability Testing Results

E-commerce Task	Results after checking the attributed existence on every e-commerce task	
	High rate Application(*Amazon*)	Low rate Application(*Waradly*)
Searching	95 %	49 %
Seeking product data	100 %	61 %
Purchasing	95 %	59%

B. Discussion

Most of the previous work papers focus on updating the needed attributes that is used in testing the applications and test them to proof their idea or taking an existence framework and upgrade the way of evaluation to get better results. But in this research, the framework focus on describing all the usability attributes from qualitative title into quantitative detailed steps that helps the developer and tester on their daily work.

VII. Limitations of the Study

For future improvement, this research contains some limitations as setting usability attributes for each application type according to the application categories. Also, differentiating between the general attributes that could be applied on all applications and the specific application types.

Finally, setting weights for each attribute to clarify the importance of each attribute and order them according to the attribute weights.

VIII. Conclusion

After testing the framework on two e-commerce applications, one with a high rate and the other with a low rate on three main tasks, which are searching, seeking product data, and purchasing, our framework proves that a high rate (Amazon) resulted in 95% in the searching task while a low rate (Waradly) resulted in 49%. Also, it proves that a high rate (Amazon) resulted in 100% in the seeking product data task while a low rate (Waradly) resulted in 61%. Finally, a high rate (Amazon) resulted in 95% in the purchasing task while a low rate (Waradly) resulted in 59%.

So that means our framework proved that Amazon has a high rate and Waradly has a low rate, which means it can measure the usability percentage and categorises the application usability. The testers can use the framework in testing their applications of any type as the attributes are generic, which means they can measure any type of application.

Acknowledgment

First and foremost, I feel always indebted to Allah the Most Beneficent and Merciful. Words can never suffice to reflect my sincere appreciation and gratitude to Prof. Dr. Hesham Hassan, Prof. Dr. Khaled Wassif, and Dr. Hanaa Bayomi.

References

[1] M. A. Umar, "Comprehensive study of software testing: Categories, levels, techniques, and types," International Journal of Advance Research, Ideas and Innovations in Technology, 2019.

[2] Ashraf Mousa Saleh and Roesnita Binti Ismail, "Usability Evaluation Frameworks Of Mobile Application: A Mini-Systematic Literature Review," in The 3rd Global Summit on Education GSE 2015At: Kuala Lumpur, MALAYSIA, 2015.

[3] ASHRAF SALEH, ROESNITA BINTI ISAMIL, NORASIKIN BINTI FABIL, "EXTENSION OF PACMAD MODEL FOR USABILITY EVALUATION METRICS USING GOAL QUESTION METRICS (GQM) APPROACH," Journal of Theoretical and Applied Information Technology, 10th September 2015.

[4] Nielsen, Usability Inspection Methods, New York: NY: John Wiley and Sons, 1994.

[5] Shabana Shareef and M.N.A. Khan, "Evaluation of Usability Dimensions of Smartphone Applications," International Journal of Advanced Computer Science and Applications, Vols. Vol. 10, No. 9, 2019.

[6] Nigel Bevan, Jim Carter, Jonathan Earthy, Thomas Geis, Susan Harker, "New ISO Standards for Usability, Usability Reports and Usability Measures," M. Kurosu (Ed.): HCI 2016, Part I, LNCS 9731, p. 268–278, July 2016.

[7] ANUBHA GULATI and SANJAY KUMAR DUBEY, CRITICAL ANALYSIS ON USABILITY EVALUATION TECHNIQUES, International Journal of Engineering Science and Technology (IJEST), 2012.

[8] Nigel Bevan, Jim Carter, Jonathan Earthy, Thomas Geis and Susan Harker, "ISO 9241-11 Revised: What Have We Learnt About Usability Since 1998?," in M. Kurosu (Ed.): Human-Computer Interaction, Part I, HCII 2015, LNCS 9169, Springer International Publishing Switzerland, 2015, p. 143–151.

[9] Bevan N. & Macleod M., "Usability measurement in context.," in Behaviour and Information Technology, 1994, p. 132–145.

[10] Bimal Aklesh Kumar & Munil Shiva Goundar & Sailesh Saras Chand, "Usability guideline for Mobile learning applications: an update," Education and Information Technologies, june 2019.

[11] Sidra Shahid and Muhammad Shabbir Abbasi, "Usability Testing of an E- Learning System:A Comparative study of two Evaluation Techniques," IOSR Journal of Computer Engineering, vol. 16, 2014.

[12] Bimal Aklesh Kumar & Munil Shiva Goundar & Sailesh Saras Chand, "Usability Guideline for Mobile Learning Applications," Education and Information Technologies, jan 2019.

[13] Estela Aparecida, Oliveira Vieira, Aleph Campos Da Silveira and Ronei Ximenes Martins, "Heuristic Evaluation on Usability of Educational Games: A Systematic Review",," Informatics in Education, Vols. Vol. 18, No. 2, p. 427–442, 2019.

[14] Bimal Aklesh Kumar and Priya Mohite, "Usability of mobile learning applications: a systematic literature review," J. Comput. Educ., 2017.

[15] Muhammad Salman Bashir and Amjad Farooq, "EUHSA:: Extending Usability Heuristics for Smartphone Application," IEEE Access, vol. 4, 2017.

[16] RUYTHER PARENTE DA COSTA, EDNA DIAS CANEDO, RAFAEL TIMÓTEO DE SOUSA, JR2 (Member, IEEE), ROBSON DE OLIVEIRA ALBUQUERQUE and LUIS JAVIER GARCÍA VILLALBA3, "Set of Usability Heuristics for Quality Assessment of Mobile Applications on Smartphones," IEEE Access, 2017.

[17] Carmelo Ardito1 & Maria De Marsico2 & Davide Gadia3 & Dario Maggiorini3 & Ilaria Mariani & Laura Ripamonti & Carmen Santoro, "Special Issue on Advances in Human-Computer Interaction," in Multimedia Tools and Applications, Springer Science+Business Media, LLC, part of Springer Nature 2019, 2019, p. 13353–13359.

[18] Anju Gautam, Prashant Sahai Saxena & Savita Shiwani, "Quantitative analysis of usability issues of Ecommerce portal," Journal of Statistics and Management Systems, pp. 681-692, 2017.

[19] P. WEICHBROTH, "Usability of Mobile Applications: A Systematic literature study," IEEE Access, 2020.

[20] Jakub Myka, Agnieszka Indyka-Piasecka, Zbigniew Telec, Bogdan Trawiński, Hien Cao Dac, "Comparative Analysis of Usability of Data Entry Design Patterns for Mobile Applications," reearch gate, jan 2019.

[21] Azham Hussain, Nor Laily Hashim, Nazib Nordin and Hatim Mohamad Tahir, "A METRIC-BASED EVALUATION MODEL FOR APPLICATIONS ON MOBILE PHONES," Journal of ICT, vol. 12, p. 55–71, 2013.

[22] Luis Cayola and José A. Macías, "Systematic guidance on usability methods in user-centered software development," Information and Software Technology, 2018.

[23] Karima Moumane, Ali Idri and Alain Abran, "Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards," springer plus, p. 15, 2016.

[24] Azham Hussain, Emmanuel O.C. Mkpojiogu, Nur Hafiza Jamaludin and Somia T.L. Moh, "A usability evaluation of Lazada mobile application," American Institute of Physics, 2017.

[25] Nouzha Harrati, Imed Bouchrika, Abdelkamel Tari and Ammar Ladjailia, "Automating the Evaluation of Usability Remotely for Web Applications via a Model-Based Approach," First International Conference on New Technologies of Information and Communication (NTIC), 2015.

[26] Ger Joyce, Mariana Lilley, Trevor Barker and Amanda Jefferies, "Heuristic Evaluation for Mobile Applications: Extending a Map of the litrature," in Advances in Usability, User Experience and Assistive Technology , 2019, pp. (pp.15-26).

[27] Christiane Gresse von Wangenheim, Talita A. Witt, Adriano Ferreti Borgatto, Juliane Vargas Nunes, Thaisa Cardoso Lacerda, Caroline Krone, Laís de Oliveira Souza., "A Usability Score for Mobile Phone application based on Heuristics," International Journal of Mobile Human Computer Interaction, 2016.

[28] Zahra Al Mahdi, Vikas Rao Naidu, and Preethy Kurian, "Analyzing the Role of Human Computer nteraction Principles for E-Learning solution design," Smart Technologies and Innovation for a Sustainable Future, Advances in Science, Technology & Innovation, 2019.

[29] Azham Hussain and Emmanuel O.C. Mkpojiogu, "Usability Evaluation Techniques in Mobile Commerce Applications: A Systematic Review," Proceedings of the International Conference on Applied Science and Technology, 2016.

[30] Mamta Pandeya, Ratnesh Litoriyaa, and Prateek Pandeya, "Application of Fuzzy DEMATEL Approach in Analyzing Mobile App Issues," Programming and Computer Software,, 2019.

[31] Renuka Nagpal, Deepti Mehrotra, Pradeep Kr. Bhatia and Arun Sharma, "FAHP Approach to Rank Educational Websites on Usability," International Journal of Computing and Digital Systems, 2015.

[32] Mamta Pandey, Ratnesh Litoriya, Prateek Pandey, "Identifying Causal Relationships in Mobile App Issues An Interval Type 2 Fuzzy DEMATEL Approach," wireless personal communication, 2019.

Design and Development of Face Mask Reminder Box Technology using Arduino Uno

Chee Ken Nee[1]*, Rafiza Abdul Razak[2], Muhamad Hariz Bin Muhamad Adnan[3]
Wan Fatin Liyana Mutalib[4], Nur Fatirah Roslee[5], Noor Mursheeda Mahyuddin[6]
Department of Computing, Faculty of Arts, Computing and Creative Industry
Sultan Idris Education University, 35900 Tanjung Malim, Perak, Malaysia[1, 3]
Curriculum & Instructional Technology Department
Faculty of Education, University of Malaya, 50603 Kuala Lumpur, Malaysia[2, 4, 5, 6]

***Abstract*—The World Health Organization (WHO) declared the COVID-19 pandemic on 12 Mar, 2020, due to the growth in the number of cases worldwide. WHO advises wearing a face mask and practicing social distancing, which has played a crucial role in prevention and control measures that can prevent the spread of COVID-19. Thus, this paper presents the process through which face mask box is equipped with a voice reminder and sensor. It is made with the help of an Arduino Uno board to give awareness or reminder whenever a person is alerted with a voice reminder to wear a face mask before going outside. It can be helpful, especially in the pandemic era, as a new norm of practice in wearing a mask.**

***Keywords*—Face mask box; COVID-19; Voice reminder; Arduino; New norm**

I. Introduction

The World Health Organization (WHO) declared COVID-19 as a pandemic on 12 Mar, 2020 due to the growth in the number of cases worldwide [1]. The Covid-19 pandemic has had a negative impact on health, education, economy, finance, and others, resulting in the introduction of the Movement Control Order for many countries. People were advised to stay at home, work from home, limit their movement outside the house, wear a mask to cover their nose and mouth, regularly wash their hands, establish social imprisonment, and so on as precautions to stop this pandemic [2]. So, all people need to carry and adapt to daily life with the new norm; as a prevention and control measure to limit the spread of COVID-19, WHO advises wearing a face mask and practicing social distancing [3]. Mask has played a crucial role in prevention and control measures to prevent the spreading of some respiratory diseases and COVID-19. The use of masks and frequent hand hygiene, social distancing, and other Infection Prevention and Control (IPC) measures should be followed to prevent the spread of COVID-19 [4].

One significant benefit of wearing face masks is to protect the people around us, especially if we have been exposed to someone carrying the virus or work in a crowded place. Therefore, wearing masks is a critical habit everyone should be applying since this simple step could significantly reduce the risk of transmission [5]. Hence, it is a pressing need to design and develop a face mask reminder with a low-cost material that can be easily made since there is also none of this product available in the market to date.

II. Background and Previous Related Work

A. Motivation and Problem Statement

Coronavirus Disease 2019 or known as COVID-19, has become the latest threat to global health [6]. It is an infectious disease [7] that causes severe acute respiratory syndrome (SARS) [6]. The symptoms of the infection consist of fever, cough, nasal congestion, fatigue, and other signs of upper respiratory tract infections [8]. The condition can cause severe disease or death; as a prevention and control measure to limit the spread of COVID-19, the WHO advises wearing a face mask and practicing social distancing [3].

Recently, it seems like some people are starting to disregard the WHO order to practice social distancing and wear a face mask whenever they leave the house or are in a crowded place. Thus, to address these problems, every country affected by COVID-19 has made it mandatory for people to wear face masks in public, especially in crowded places. The new normal must be continued depending on our awareness level, attitudes, and habits. For this project, the problem focuses on the attitude and practices of wearing face masks. Thus, to solve the issues stated, this project developed a face mask reminder to give awareness of the importance of face masks. A person will be alerted with a voice reminder to wear a face mask before going outside through the project.

For this project, the problem focuses on the attitude and habits of wearing face masks. Thus, to solve the issues stated, this project developed a face mask reminder to give awareness of the importance of face masks. A person will be alerted with a voice reminder to wear a face mask before going outside through the project.

B. Previous Related Work

Based on the literature review of the previous works, the ideas of face mask boxes and voice reminders are considered to develop an awareness system toward the importance of face masks. The difference from previous related work is through the project, a person will be alerted with a voice reminder to wear a face mask before going outside.

*Corresponding Author

Research Management and Innovation Center (RMIC), Universiti Pendidikan Sultan Idris.

1) Build your own COVID-19 face mask reminder using Arduino: The project designed and made by The Assembly Production addressed how people always forgot to wear a face mask when going outside [9]. The emergence of COVID-19 has made the entire world obey the rules on social distancing and wear face masks to prevent the spread of COVID-19.

Thus, the Assembly Production designed a face mask reminder using Arduino Uno, PIR sensor, buzzer, LCD, and keypad. The PIR sensor detects the motion of a person going through the device, and the buzzer will make a noise to alert the person to wear a face mask before going outside [9]. The device also has an LCD to display the number of face masks and a keypad to enter the number of face masks in the box. Fig. 1 shows the device produced by the Assembly where the user can store the face mask in the designated container. The face mask's material was a paper box, which was low in cost.

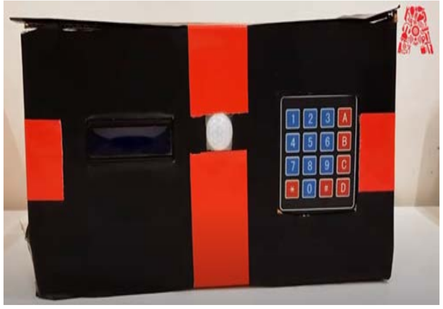


Fig. 1. Face Mask Reminder Device.

2) Contact-free hand wash dispenser with voice assistance: The Assembly Production The project designed and made by Neutrino addressed how to use a hand wash without making any contact while taking the hand soap, as illustrated in Fig. 2 [10]. It is designed using Arduino Pro Mini, DF Player, PIR sensor, DS1307 Tiny RTC, SD Card module, and speaker. The intelligent hand wash dispenser reminds a user with a message after every 2 hours to wash hands. The PIR sensor detects the motion of a person by putting their hands next to the hand soap. Then, the signal is sent to the DF player, and voice assistance will help them wash their hands.

Fig. 2. Contact - Free Hand Wash Dispenser.

III. Development of Face Mask Reminder Prototype

A. Methods

The face mask reminder is designed to alert the person to wear a face mask before going out. Fig. 3 shows the flowchart for project planning.

To assure the project runs smoothly, systematic action was taken to analyze the face mask reminder's performance. The project flow is divided into four phases: Research, Design and Modelling, Prototype, and Result and Analysis.

For Phase 1: Research, the project begins with a detailed study on a literature review about voice reminders and motion detection. Based on the previous related works, [9] designed a face mask reminder box to noise to alert the person to wear a face mask before going outside. While [10] developed a contact-free hand wash dispenser with a voice reminder feature. The project will also focus on the simulation and coding to implement motion detection and voice reminder.

For Phase 2: Design and Modelling, the project followed by designing the device for a better experience. In this phase, some research was done on hardware and software compatibility and suitability for the proposed system. Every system is a combination of hardware and software, mainly hardware or software [11].

For Phase 3: Prototype, the prototype's functionality is tested after the simulation and design is done. The design was done using Fritzing to determine the hardware connection of each component. At the same time, the simulation of hardware connection was done by using Proteus software.

For Phase 4: Result and Analysis, the result obtained from testing the prototype was analyzed to monitor the system's performance. The analysis was done on the device's performance and improvement on other parts.

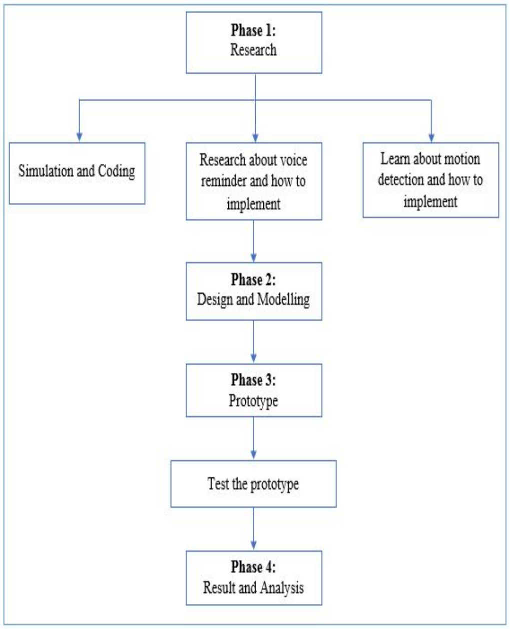


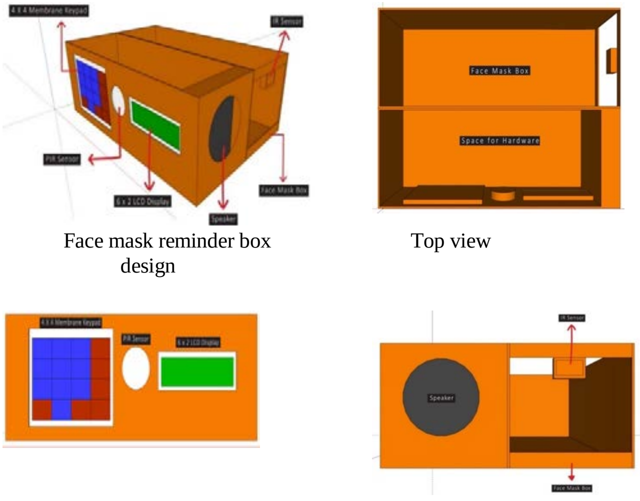
Fig. 3. Flowchart of the Project Planning.

B. Face Mask Reminder Box Design

To assure the project runs smoothly, systematic action was taken to analyze the face mask reminder's performance. The project flow is divided into four phases: Research; the Face Mask Reminder is designed on Sketch Up software based on the requirements.

The prototype design includes space for hardware and a face mask box. Fig. 4 shows the face mask reminder box's design using SketchUp software. SketchUp has a lot of possibilities and needs a low amount of pre-knowledge to get started. This software shows some knowledge about models, building and using models can be getting through its help. There might be a few doubts for the new users, but this can be solved easily with help and demonstrated videos. There are tools for selection, drawing, component, view, and sharing the drawing designs [12]. The sketching shows the sides to be used to place all the hardware with each part labeling. The 4x4 keypad membrane, PIR sensor, and 16x2 LCD display are placed side by side for the front side. For the right side, the speaker and IR sensor are also placed side to side.

Hardware space is for placing the Arduino UNO and power supply; the suggested power supply to be used is a power bank. Arduino is used as an open-source computing platform that is used for constructing and programming electronic devices. It can also act as a mini-computer, like other microcontrollers, by taking inputs and controlling the outputs for various electronic devices [13]. At the same time, the face mask box is for the user to put the face mask in the box up to 100 pieces of face mask. The IR sensor function will detect a person's hand that takes out the face mask, and the number displayed on the LCD will be reduced. Then, the PIR sensor will detect a person's movement through the box and play a voice reminder: "Please wear a face mask before going out" using a speaker. So people will become more alert during going out.



Face mask reminder box design

Top view

Front view

Right Side View

Fig. 4. Design of Face Mask Reminder Prototype using SketchUp Software.

C. Mask Reminder Box Fabrication

The face mask reminder is designed to alert the person to wear a face mask before going out. As shown in Fig. 4, the face mask reminder box design is divided into two parts. The first part is for hardware connection space, where all the hardware will be placed in this space. This box comes with a small hole for power supply purposes. The second part is for a face mask box where a user can store up to 100 face masks in this space.

The material used for the box is Acrylic or Perspex. Acrylic is a transparent plastic material with outstanding strength, stiffness, and optical clarity. Acrylic is also a material with properties such as transparency and durability. It is now being used in a wide range of applications such as lenses of glasses, tail lights, and various other instruments in a vehicle to reduce cost and productivity [14]. It is highly resistant to variations in temperature and humidity. Thus, it is helpful in outdoor applications as well. Fig. 5 shows the top view of the box. The box has a lid that can be opened to check the hardware connection and fill in the face mask. The cover can be opened and closed quickly.

Fig. 6 shows the front view of the face reminder box. The 4x4 membrane keypad and 16x2 LCD keypad with I2C are placed side to side on the front side. The keypad is used to key in the number of face masks in the box. The LCD display presents the actual number of face masks in the box. On display, it will show the exact number of masks and the maximum number that can be entered. The maximum number of face masks that can be placed in the box is 100 pieces. The device can be used easily if the hardware is connected.

Fig. 7 shows the left side view of the face reminder box. The speakers, PIR, and IR sensor are placed side to side on the left side. In addition, on the right is a face mask box where the user can place a face mask. The speaker is used to alert "Please wear a face mask before going out" to anyone passing through the box. PIR sensor is used to detect a person's motion. When it detects a motion, the speaker will sound an alert. At the same time, an IR sensor is used to detect a hand motion of a person. If a person removes a face mask from the box, the number displayed on the LCD will be reduced.

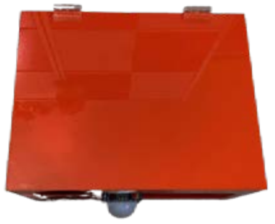


Fig. 5. Top View of Face Mask Reminder Box.

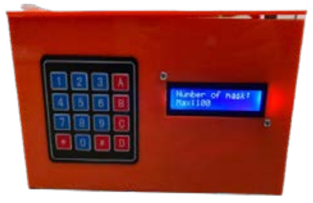


Fig. 6. Front View of Face Mask Reminder Box.

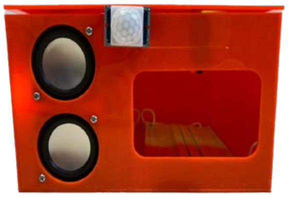


Fig. 7. Left Side View of Face Mask Reminder Box.

Fig. 8. Right Side View of Face Mask Reminder Box.

Fig. 8 shows the right-side view of the face reminder box. There is only a tiny cutout of acrylic sheet for connecting the hardware to the power supply on the right side. But, since the power supply is a small 20000mAH power bank, it is just placed inside the box. Fig. 9 shows the backside view of the face mask reminder box.

Fig. 9. Backside View of Face Mask Reminder Box.

IV. Development of Face Mask Reminder System

A. *Circuit Design of Face Mask Reminder*

The hardware component for the prototype is Arduino UNO with PIR and IR sensor, LCD with I2C, Keypad, SD card module, and Speaker. Fig. 10 shows the circuit connection using Fritzing application.

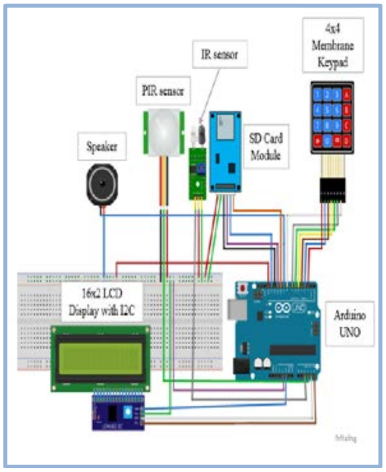


Fig. 10. Hardware Connection of Face Mask Reminder using Fritzing.

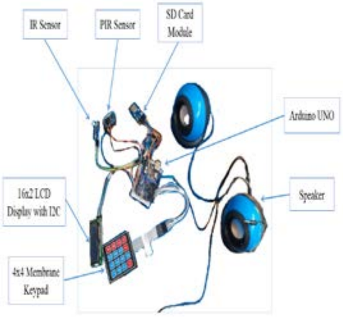


Fig. 11. Hardware Connection of Face Mask Reminder.

The prototype consists of only hardware parts tested to assure the device's functionality. Fig. 11 shows the hardware connection of the face mask reminder. Arduino UNO acts as a microcontroller and holds all of the other components. The power supply is a 20000mAH power bank to keep the hardware active.

V. Implementation of Face Mask Reminder

A. *Testing the Face Mask Reminder Box*

The face mask reminder box was tested to get its functionality in this part. The test was done in a living room near a door. A hardware component is connected to the 20000mAH power bank that acts as a power supply to keep the box active through the testing. The audio of different voice alerts is stored on an SD card in waveform audio file (WAV) format [15].

A person passed through the box, and the speaker sent an alert with a voice reminder, "Please wear a face mask before going out." Next, the box reminds the person to wear a face mask before going outside to prevent exposure to the virus COVID-19. Then, the person takes out a face mask from the box and decreases the number of face masks shown on the LCD display.

On the other hand, the person refills the face mask in the box. After that, he keys in the actual amount of face mask added using the keypad. Thus, the exact number of face masks in a box appeared on the LCD.

The result obtained in this project is analyzed according to the related part. It has been proved that the face mask reminder box works systematically according to the working principle and based on the testing phase. As a result, the face mask reminder is successfully implemented.

B. *Manual to Key in Quantity of Face Mask*

Fig. 12 shows that the keypad matrix with 16 push buttons [16] is used to key in the number of face masks in a box. This type of keypad has four rows and four columns where the overlapping rows and columns are the keys [17]. To key in the number of face masks, the user can follow the instruction below.

1) Press any number from 1 to 100.
2) Press "#" to enter.
3) The number will be shown on the LCD.
4) To backspace, press "A."

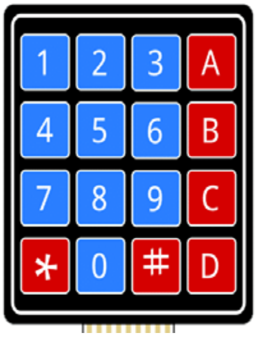


Fig. 12. 4x4 Keypad Membrane.

Users can only enter numbers 1 to 100 because the system only allowed face masks up to 100 pieces. If a user enters more than 100, the LCD will show "Invalid number," and it must key in the number again. If the user enters a wrong number, press keypad "A" during the process to backspace.

VI. Overall System Architecture

The overall system architecture of the developed Face Mask Reminder system is illustrated in Fig. 11 and Fig. 13. The system design consists of one central part: hardware. The hardware components are Arduino UNO, PIR and IR sensor, 16x2 LCD with I2C, 4x4 membrane keypad, SD card module, and speaker.

Firstly, Arduino UNO acts as a microcontroller that integrates with the PIR sensor to detect any motion within the range [11]. When the sensor detects a person's movement, a speaker will alert with a voice reminder from the SD card module. It will be a warning for a person to wear a face mask before going outside. Secondly, Arduino UNO is a microcontroller that integrates with an IR sensor to detect the hand motion of someone taking a face mask from the box [18]. When the sensor detects a hand motion, the number displayed on the LCD will be decreased. Lastly, the keypad is used to enter the number of face masks in the box.

The flowchart of the system operation mechanism is shown in Fig. 14. The application starts with the face mask reminder box attached next to the door or somewhere close to the door. Then, a user can fill in the face mask in the box and key in the quantity of the face mask using the keypad membrane. The actual number of face masks in a box is shown on the LCD.

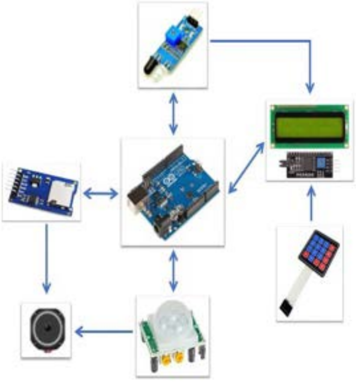


Fig. 13. Face Mask Reminder System Architecture.

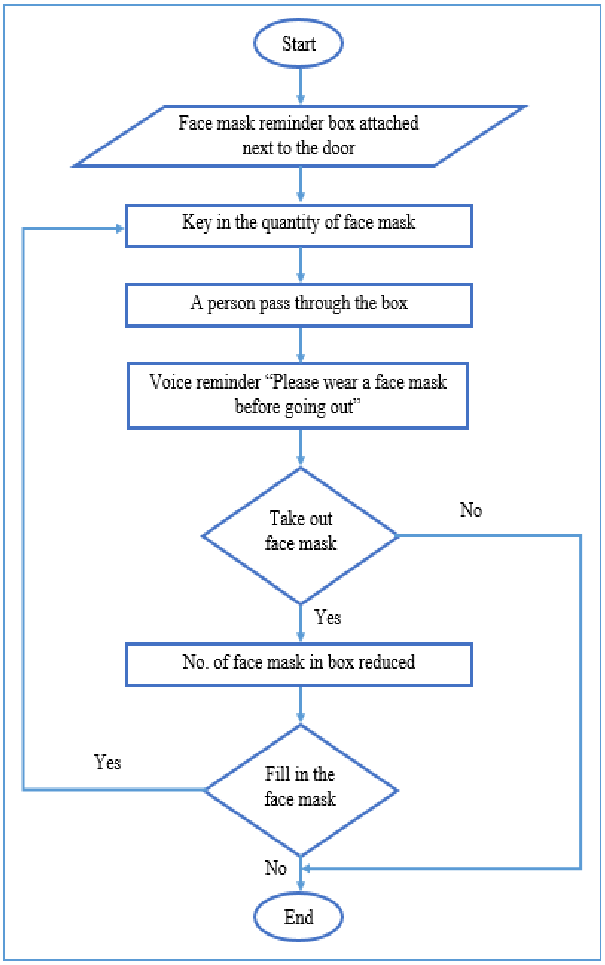

Start
Face mask reminder box attached next to the door
Key in the quantity of face mask
A person pass through the box
Voice reminder "Please wear a face mask before going out"
Take out face mask
No
Yes
No. of face mask in box reduced
Yes
Fill in the face mask
No
End

Fig. 14. Flowchart of Face Mask Reminder Application.

When a person passes through the box, the PIR sensor detects a motion and sends a signal to the speaker [19]. A voice reminder alerts the person with a warning, "Please wear a face mask before going out." If a person takes a face mask from the box, the IR sensor senses a hand motion into the face mask box [20]. Then, the number of face masks in the box was reduced. Thus, if a user fills in the face mask in the box, they must key in quantity, and the actual number appears on the LCD [21]. If a person does not make a face mask or fill in the face mask, they can continue their way out. The complete supplementary files for the Facemask Reminder coding algorithm can be found in the link here: https://tinyurl.com/facemaskbox reminder

VII. Conclusion and Recommendation

The design of the proposed system only consists of hardware parts. The hardware implementation used Arduino UNO, PIR and IR sensor, LCD Display, Keypad, Speaker, and SD Card Module. It is used to develop the Face Mask Reminder, which provides a face mask box with a voice reminder. The Face Mask Reminder box capabilities provide a voice reminder as an alert or warning to a person to wear a face mask before going outside. To reduce the problem of people who recently forgot to wear a face mask, the box comes with a space to place a face mask up to 100pcs. The sensors increased accuracy in detecting a person's motion when it passed through the box and taking a face mask out of the box. The performance of the box is tested in the living room, and it shows an excellent result that can be used for a face mask reminder. The Face Mask Reminder box is built successfully. However, this system could be upgraded better by using more accurate sensors. PIR sensors depend on motion detection, with the motion sensitivity declining as the user's distance from the sensor increases [22]. While IR sensor is highly accurate, they are highly fragile in nature [23].

ACKNOWLEDGMENT

This research was supported by Universiti Pendidikan Sultan Idris (UPSI) through funding from Research Management and Innovation Center (RMIC) for the publication fee in this journal.

REFERENCES

[1] Ministry of Health. (2020). COVID-19: Social Distancing Guidelines for Workplace, Homes and Individuals. Guidelines COVID-19 Management. Annex 26. No. 5/2020. Retrieved from http://covid-19.moh.gov.my/garis-panduan/garis-panduan-kkm.

[2] World Health Organization (2020). Coronavirus disease (COVID-19) advice for public: When and how to use masks. Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-usemasks?adgroupsurvey={adgroupsurvey}&gclid=Cj0KCQjwpdqDBhCSARIsAEUJ0hNx_wVZvjBA43M9PfAqjWaH51d_IVzI7XOcmAUD7VKhseupyH6_p8waAr65EALw_wcB.

[3] World Health Organization. (2020). Mask use in the context of COVID-19: interim guidance. COVID-19: Infection prevention and control. Emergencies Preparedness, WHO Headquarter (HQ). WHO/2019-nCoV/IPC_Masks/2020.5 Retrieved from https://www.who.int/publications/i/item/advice-on-the-use-of-masks-in-the-community-during-home-care-and-in-healthcare-settings-in-the-context-of-the-novel-coronavirus-(2019-ncov)-outbreak.

[4] Ministry of Health. (2020). Annex 8a: Guidance on the Use of Masks with Regards to Covid-19 Pandemic. Guidelines COVID-19 Management. Annex 8. No. 5/2020. Retrieved from http://covid-19.moh.gov.my/garis-panduan/garis-panduan-kkm.

[5] World Medical Card (2020). The Importance of Wearing Masks. Retrieved from https://www.wmc-card.com/uk/the-importance-of-wearing-masks/.

[6] Fauci, A. S., Lane, H. C., & Redfield, R. R. (2020). Covid-19 — Navigating the Uncharted. New England Journal of Medicine, 382(13), 1268–1269. https://doi.org/10.1056/nejme2002387.

[7] Post-COVID-19 global health strategies: the need for an interdisciplinary approach. (2020). Aging Clinical and Experimental Research, 32(8), 1613–1620. https://doi.org/10.1007/s40520-020-01616-x.

[8] Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 Epidemic. Tropical Medicine & International Health, 25(3), 278–280. https://doi.org/10.1111/tmi.13383.

[9] The Assembly Production. (2020). Build Your Own COVID-19 Face Mask Reminder Using Arduino. The Assembly Production Team. Retrieved from https://youtu.be/yu8g9jn41zU.

[10] Neutrino. (2020). Contact-Free Hand Wash Dispenser with Voice Assistance. Neutrino Team. Retrieved from https://youtu.be/N_-7oMFJ18k.

[11] Budijono, S., Andrianto, J., & Axis Novradin Noor, M. (2014). Design and implementation of modular home security system with short messaging system. EPJ Web of Conferences, 68, 00025. https://doi.org/10.1051/epjconf/20146800025.

[12] Dr.Pradip Ashok Symote (2016). Google Sketch up: A Powerful Tool for 3D Mapping and Modeling. International Journal of Computer Application and Engineering Technology. Volume 5-Issue. Pp. 377 - 382.

[13] Leo Louis (2016). Working Principle of Arduino and using it as a Tool for Study and Research. Department of Electronics and Communication Engineering, Gujarat Technological University, Ahmedabad, India. International Journal of Control, Automation, Communication and Systems (IJCACS), Vol.1, No.2.

[14] Eshwar Pawar (2016). A Review Article on Acrylic PMMA, IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE) e-ISSN: 2278-1684, p-ISSN: 2320-334X, Volume 13, Issue 2 Ver. I.

[15] Sharmad, P. (2016). Thingspeak Based Sensing and Monitoring System for IoT with Matlab Analysis. International Journal of New Technology and Research, 2(6), 19-23. https://www.ijntr.org.

[16] Arduino Website. Accessed: Nov. 5, 2020. [Online]. Available: https://www.arduino.cc/en/software.

[17] Khan, S., Rahman, O., & Ehsan, M. (2017). Design and Fabrication of a Password Protected Vehicle Security and Performance Monitoring System. 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 568–571. https://doi.org/10.1109/r10-htc.2017.8289022.

[18] Rindhe, C. (2020). Smart Car Parking System using IR Sensor. International Journal for Research in Applied Science and Engineering Technology, 8(5), 2460–2462. https://doi.org/10.22214/ijraset.2020.5405.

[19] Woodstock, T. K., & Karlicek, R. F. (2020). RGB Color Sensors for Occupant Detection: An Alternative to PIR Sensors. IEEE Sensors Journal, 20(20), 12364–12373. https://doi.org/10.1109/jsen.2020.3000170.

[20] Kumar, P. L. (2020). Vehicle Speed Detection System using IR Sensor. International Journal for Research in Applied Science and Engineering Technology, 8(5), 1563–1567. https://doi.org/10.22214/ijraset.2020.5253.

[21] Rajkumar, M., Anamika, & Sushmita, K. (2018). Biometric Based Electronic Voting Machine. International Journal of Engineering Development and Research, 6(2), 273-277.

[22] Lopez-Belmonte, J., Marin-Marin, J. A., Soler-Costa, R., & Moreno-Guerrero, A. J. (2020). Arduino Advances in Web of Science. A Scientific Mapping of Literary Production. IEEE Access, 8, 128674–128682. https://doi.org/10.1109/access.2020.3008572.

[23] Kumar, P. L. (2020). Vehicle Speed Detection System using IR Sensor. International Journal for Research in Applied Science and Engineering Technology, 8(5), 1563–1567. https://doi.org/10.22214/ijraset.2020.5253.

An Efficient Unusual Event Tracking in Video Sequence using Block Shift Feature Algorithm

Karanam Sunil Kumar[1]
Assistant Professor
Department of Computer Science and Engineering
RNS Institute of Technology
Bangalore, India

Dr. N P Kavya[2]
Professor
Department of Computer Science and Engineering
RNS Institute of Technology
Bangalore, India

***Abstract*—The area of video technology is rapidly growing owing to advancements in intelligent video systems in sensor operations, higher bandwidth capacity, storage, and high-resolution displays. This led to the proliferation of video-based computing modeling to perform specific tasks on video sequences to gain more insight from the data. Visual tracking of events is a core component in video visual surveillance systems that classify and track moving objects to describe their behavioral aspects. The prime motive behind intelligent video systems is to perform efficient video analytics to meet the specific requirements of the user/use-cases. It involves a self-directed paradigm to understand event sequences, reducing the computational burden of characterizing the activities. The study incorporates a block-shift feature algorithm and introduces a novel computational research method for unusual event tracking in video sequences. The formulated approach employs a framework combining operational blocks to compute sequential operations such as block-matching from the dictionary of motion estimations. Before applying the learning model, the subsequent analysis procedure adds feature lexicon and dominant attributes to make the execution computationally efficient. Further, it uses a sparse-non negative factorization approach to organize the informative details into k possible finite clusters. The event detection outcome from the training datasets of video sequences shows better experimental results than the traditional highly cited related approach of unusual object detection and tracking.**

Keywords—Object detection; tracking; learning models; video sequence analysis

I. Introduction

The ever-increasing surveillance process adopted in every walk of life leads to the video sequence databases' reposit. If these databases are not analyzed, it is just a kind of dead data. The analysis of textual data is quite simple as the basic terms construct it. In contrast, the video sequences contain various sensitive pattern features and information. Its interpretation requires suitable and efficient algorithms to deal with its analysis so that multiple applications can be built based on this analysis, and real-time decisions can be taken. The advancement of the computing platforms with the CPU and GPU functioning provides the right ecosystem to explore the possibilities of automatically tracking an object in the video sequences. In the computer vision domain, detecting and tracking the objects within video sequences plays an essential role in many applications [1]. Some of the examples of such applications include i) event tracking in video surveillance [2], ii) dynamic road traffic management [3], iii) intruder tracking [4], and iv) robotics vision [5]. The individual image sequences in the video are popularly known as a frame in video processing. These frames contain both the static and moving object into it. The structure's static part is termed the background, whereas the moving objects are known as the foreground [6]. The task of tracking objects in the video frames or sequences includes detecting moving objects of interest, then classifying them and following them from frame to frame.

In the process of object detection, the pixels of the object in interest are clustered, which is generally performed by various methods, including i) frame differences, ii) subtraction of the background, and iii) optical flow computation [7]. The object classification occurs only after the object detection based on the high-level features, including one or combinations of color, shape, texture, or sometimes motion. However, the tracking of an object is to gain specific information on the object utilizing its orientation, activity, occlusion, etc., using three popular approaches i) silhouette [8], ii) kernel [9], and iii) point-based tracking [10]. Point-Based Tracking (PBT) is significant for tracking small objects in the video sequences; however, it does not perform well in occlusion and provides false detection [11]. The PBT is broadly classified into i) Kalman filtering (KF) [12] and ii) Particle filtering (PF) [13]. The typical working process of the KF is to perform prediction of the current state variables and correction recursively. Since KF handles the noisy input data in recursion, it can track single objects in a real-time scenario [14]. Though the PF also performs the prediction and correction similar to the KF, it overcomes the state variable's approximation constraints by re-sampling.

PF generates many possible models for the current variable before moving to the following state variable. The PF exploits feature sets like {color, contour, texture}[15]. The Kernel-Based Tracking (KBT) uses shapes like rectangles or ellipses to keep the object of interest outside the shape and background to detect rigid objects. The typical classification of the KBT includes i) Simple Template Matching (STM), ii) Mean Shift (MS), iii) Support Vector Machine (SVM), and iv) Layer Matching (LM) [16]. The STM is applicable for tracking a single object in partial occlusion conditions. It verifies the video sequences with a reference frame, and only motion transformation is possible in this method. Simultaneously, the MS method uses chamfer distance transformation, the Bhattacharya coefficient for distance, and color distribution transformation among the region of interest in windows to windows to improve the accuracy. However, it cannot track high-speed moving objects and only single object tracking [17]. However, the SVM-based method tracking the object in video sequences tracks only the positive samples. This method also handles tracking a single object and limits its applicability to partial occlusion [18]. The LM- method is used along with the KB method to track multiple objects, even in the complete occlusion condition [19]. The complex object containing a composite shape is well followed using the Silhouette-based method, which is further categorized into i) contour tracking (CT) [20] and ii) shape matching (SM) [21]. In CT, the object's motion and shape are considered using a state-space model. Then, the contour energy is minimized using gradient descent for computing the next frame's contour in iteration from the previous frames. This method helps track objects with irregular shapes. However, the shape-based process is the same as SMM without classifying the objects, and it uses the edge templates and occlusion utilizing Hough transformations.

This paper proposes a moving object tracking with satisfying background conditions using a block-matching method from the dictionary of the motion estimations. The contribution and potential benefits of proposed approach are as follows:

- The presented model introduces a learning-based approach, initially incorporates-video sequence exploration, followed by block-wise feature lexicon extraction, updates the dictionary of feature lexicon with dominant attributes.
- The system also includes a novel feature engineering schema and workflow modeling to compute feature elements with lexicon vector.
- The study model discusses about a novel sparse non-negative factorization (S-NNF) technique that factorizes data organization into k possible finite clusters.
- The study further applies a simplified learning-based systematic approach to faster tracking moving objects from the video feed of frame sequence.
- The system is analytically represented with numerical analysis, which subjects to the dictionary-based learning operation towards tracking the moving object anomaly.
- The study also simulates the devised numerical models with the systematic workflow execution for evaluating the performance. The outcome clearly shows the effectiveness of the block-shift feature algorithm in terms of different classification parameters.

All the above mentioned contribution are implemented in a sequential way. The organization of the paper is as follows: Section II discusses about existing methodologies while its identified problems are briefed in Section III. Section IV discusses about the system model while result discussion is carried out in Section V. The summary of the paper in form of conclusion is briefed in Section VI

II. Review of Literature

Various strategies are evolving to improve the video tracking system's operation and performance. To develop an effective video tracking system, the existing approaches consider selecting multiple parameters, e.g., points, primitive geometric shape, contour, Silhouette of an object, articulated shape, and skeletal models. Such conventional approaches are usually modeled using standard features, e.g., color, edges, optical flow, and texture. However, the video tracking system has recently introduced various unique techniques. The existing approach offers importance to extracting the semantic factor associated with capturing essential features for performing the track. Boukhers et al. [22] have discussed a probability-based model for obtaining trajectories of a three-dimensional object from two-dimensional video feeds. The model can estimate the object depth from the calculated focal length. The technique also uses a particle filtering method based on the Markov chain Monte Carlo process to stabilize the video better feeds with reduced time consumption for detecting an object. However, the majority of the video tracking application emphasizes more on human as an object. The challenge in this field is to differentiate the object from the background based on appearance and color. This issue is sorted out by Damotharasamy [23] by using a sparse representation that is not affected by any variations caused due to illumination. Apart from this, the issue associated with occlusion is addressed using subspace learning that extracts the visual features during the dictionary's updating process.

Further work also performs sophisticated video tracking of moving objects and recognition of specific actions. Studies using a similar methodology were also encountered in the literature using the discrete tracking mechanism by Kong et al. [24]. This approach has considered the localization of coordinates of a moving object using compressive tracking. The study has contributed to refining the scaling factor and recovering the occlusion issue. A convolution network is used in this study for recognizing the action of the moving object based on pre-defined information of the actions and extracts potential features considering the static data of an object.

Existing studies have also witnessed a video tracking system from multi-feeds that offers more elaborated information of the same scene event captured from multiple cameras mounted in different locations. A similar concept has been modeled by Lee et al. [25], where segmentation and tracking multiple objects using dual forms of the feature are applied. Feedback with multi-kernel is used for detecting local objects while performing a video track with a single camera. At the same time, the contextual information and appearance are integrated to carry out multi-camera tracking. The study also used an unsupervised learning method to improve the scalability factor. However, the study is all about tracking a single object. A unique variant of this approach was seen in Liu et al. [26], capable of tracking multiple objects. Independent from any specific object model, this approach can differentiate two similar objects based on the trajectory generation. The system uses a neural network and graphical model for tracking with a correlation between the targets. Another part of the literature on video tracking has been carried out considering detecting abnormalities present within it.

Adopting the Gaussian mixture is proven to improve the video tracking system effectively, as seen in Ratre and Panchapakesan [27]. The model offers a better classification of an object and the decomposition concept's usage using tucker tension. The study also uses cosine similarity to compare the attributes of decomposition considering the mobility-based feature, i.e., speed and shape of an object, to track the event from the video feed. Existing studies also include the text as another representation of an object from the web video for exhibiting the video tracking mechanism. Tian et al. [28] further works in this direction and come up with a Bayesian-based algorithm model for object detection and tracking from a complex video form. A similar kind of tracking of the object is also carried out by Yang et al. [29] by addressing its multi-orientation. The study uses multiple frames using dynamic programming for performing video tracking over various scenes.

Literature has also witnessed the implication of the identification and classification of a different number of objects for a video feed, as seen in the work of Wong et al. [30]. The author has developed a model independent of any apriori information of the object to perform tracking. At the same time, the classification is carried out using the learning approach of the neural network. There are specific unique categories of studies on video tracking. Existing literature has also seen the usage of super-pixels to extract fluctuation of an object's appearance with the monitoring, as seen in the work of Cheng et al. [31]. Using a deep residual network, the classification performance is improved using a correlation filter in tracking. The concept of facial recognition for automated identification of a specific human as an object is presented by Khan et al. [32]. The study also uses location information and time to carry out the tracking process. The adoption of infrared modalities and Red, Green Blue content (RGB) was used to develop a video tracking system for industrial surveillance systems, as explored in Lan et al. [33]. Adopting machine learning has addressed the issues associated with modalities' discrepancies. Another unique implementation is carried out by Liu et al. [34] and Benthem et al. [35]. A stochastic nature grammar is used over a decomposed graph to manage different attributes of the signal feeds. A learning-based approach is delivered for training this grammar model. This work addresses various possibilities of error in object recognition from the video feed. Therefore, multiple methods have been evolved to brief the associated issues in the next section.

III. Research Problem

A review of existing approaches towards improving video tracking performance shows various procedures in recent times and also outlines their limitation factors. However, a closer look into existing systems shows its emphasis on identifying a single mobile object. The studies also offer minor inclusion of contextual attributes connected with objects' mobility patterns. Various studies using trajectories are not meant to evaluate the dynamic environment over the scene. This could be a more significant challenge when tracking an object with dynamic mobility patterns in the crowd. Irrespective of approaches using the extraction of particular objects, the studies lack the consideration of other similar objects moving along with the target object, which could generate a significant number of outliers in its detection process. Although this challenge is somewhat solved using a machine learning-based approach, it should be noted that such systems are computationally complex when it relates to their practical operation of video tracking. Simultaneously, dependencies of trained feeds will also include many resources to store and process them. Such a phenomenon will lag in the bounding box's appearances over the tracked target presence within a scene. Therefore, there is a potential need for research to be carried out to extract contextual information about the target object from the video feed scene to improve accuracy. Learning-based approaches are a potential solution to overcome this challenge; however, such methods also require the smart amendment to balance the spontaneity in tracking performance and computational complexity.

IV. System Model

This study continues our prior works in [36], [37]. The formulated system model consists of various sub-computational units as i) video sequence explorer: where the training and testing data are visualized to get an intuition about the scene, ii) Dictionary feature lexicon blocks: where the dictionary based on the blocks are made, iii) Feature engineering modeling with feature element and lexicon vector computation, followed by iv) Design and development of a cost-effective learning model to estimate learning-based features which help in identifying the unusual moving object from each frame of the input video sequence and v) Exploration of the numerical outcome to justify the performance of the proposed modeling. The consecutive sections illustrate the rational description corresponding to the system modeling concept with mathematical notions.

A. Video Sequence Exploration

The system model provision to select the reposit location (R_l) of the training dataset (D_t), which consists of 'n' video sequences (Vs). The explicit function $f_1([R_l, D_t] \rightarrow$*:* Sc[Vs.(N)], the typical elements of the structure set Sc={Na, D, S}, where Na = video sequence name, D=date of creation that may trace the event's date in the surveillance system, and **S**= memory space. The statistical description of the dataset is given in Table I.

The dataset typically consists of '44' independent folders consisting of '8800' video sequences in totality, divided into 77.28 % as training data and 22.72 % as testing data. Few random video sequences from both training data and the testing data are shown below in Table II.

TABLE I. Statistics of the Dataset

Sl. No	Data Description	No of the video sequences
1	Training Data	6800
2	Testing Data	2000

B. Dictionary Feature Lexicon Block

The computational block for obtaining the 'Dictionary of Feature Lexicon' (DFL) takes three input sets {D_t, R_l, Sc[Vs.(n)]}, the detailed operations are given in the algorithm -1.

```
Algorithm 1: Block wise Feature Lexicon
Input: $D_t$ , $R_l$ , Sc[Vs(n)]
Output:DFL
Process
Start
Initiate: BS← [ n x n x n ]
        F1←Sc[Vs(n)], n=1
        B←Convert F1 into a column of block n x n
        Create Dictionary Size:
                nR←$n^3$ and nC← q [no. of columns(B)]
                oBS← $n^2$
                Cs← with Size [ $n^3$,q , D= $\frac{\lfloor N(Vs)_t \rfloor}{n}$ ]
Compute: Mean of B as $\vec{M}$
$V = \frac{B_j - \vec{M}}{\|B_j - \vec{M}\|}$

Update DFL with V
End
```

The computing model initiates a block size (BS) of [n x n x n], and the entire sequence (Vs) is arranged in columns of BS with the size n x n, as shown below:

$$\begin{bmatrix} p_{1,1} & p_{1,k} & p_{1,n} \\ \cdots & \cdots & \cdots \\ p_{m,1} & p_{m,k} & p_{m,n} \end{bmatrix} \rightarrow \text{f(BS):[B1,B2,B3,....Bq]} \rightarrow [\text{B}]_{1 \text{ x } q}$$

The vector size for storing the lexicon blocks in the dictionary is n^3 x q. The observation block size (oBS) is defined as [n x n]. The container size (Cs) for the dictionary feature lexicon is Cs [n^3, q, D], where D is the dimension as in equation (1) if D n ≤ 0.49 and as in equation (2) if D ≥ 0.49

$$D = \frac{\lfloor N(Vs)_t \rfloor}{n} \tag{1}$$

$$D = \frac{\lceil N(Vs)_t \rceil}{n} \ldots \tag{2}$$

For∀ Vs.∈ Sc, an empty vector $\vec{f}$ of the number of rows and columns as of Vs. The dimension of 'n' is created, and the pixels of ∀ Vs.∈ Sc is stored in all the nul1 matrices of $\vec{f}$. The Lexicon block dictionary for ∀ Vs.∈ Sc as a null matrix of size [n^3, q,]. The mean of the B is computed as $\vec{M}$ and further, the normalized Vector (V) is computed as in equation (3).

$$V = \frac{B_j - \vec{M}}{\|B_j - \vec{M}\|} \ldots \tag{3}$$

The dictionary for the feature lexicon gets updated with the corresponding values of the V.

C. Feature Engineering

The multidimensional array for the Lexicon of 'Vs' dictionary as $\vec{\boldsymbol{L}}$(m x n x d) is an input vector for the feature engineering process. The 'Fs' A structure for storing the feature vectors ($\overrightarrow{\boldsymbol{Fv}}$).

```
Algorithm-2: Feature Engineering
Input: $\vec{L}$
Output: $\overrightarrow{Fv}$
Start
for∀ n ∈ $\vec{L}$, compute $\overrightarrow{Lt}$(m,1, d)
        initiate, X[0](m x 1)
        for ∀ d ∈ $\overrightarrow{Lt}$
                [val](m x d) ←$\overrightarrow{Lt}$(d)
    end
check for noise (NAN)
        $\vec{X}$ [0:NAN]←NAN: (Algorithm-3)
Define the number of clusters: k
        Invoke, f $_{Non-NLS}$ (): (Algorithm-4)
        {features}←NNLS($\vec{X}$,k)
Updates, $\overrightarrow{Fv}$({features}
End
```

This phase of the study incorporates an efficient feature engineering modeling to normalize the lexicon attributes $\vec{L}$(m x n x d) ∈ $'$Vs.$'$ dictionary. The prime underlying motive of the feature engineering modeling in this research phase is to speed up the calculations during the execution phase of the empirical decomposition model. A matrix structure 'Fs' is further created to update the trainable extracted features for the feature vectors ($\overrightarrow{Fv}$). The algorithm finally yields a vector of computed ($\overrightarrow{Fv}$). The computation steps exhibit that it initially computes and creates a structure corresponding to the lexicon attributes of the dictionary feature set. Here, for each lexicon attribute of the dictionary feature matrix, the process computes each column corresponding to $\vec{L}$ Moreover, further computed in a decomposed form as: $\overrightarrow{Lt}$(m,1,d) ∈ $\vec{L}$(m x n x d). The matrix decomposition process makes the computational process efficient from the execution and memory "S" viewpoint. Here the computation takes place with one column vector from the matrix. The analytical algorithm also initializes another form of the matrix: X[0](m x 1), and for each individual d, the process workflow computes dominant and significant attributes of the Lexicon from $\overrightarrow{Lt}$(d) and further store it into [val](m x d). The dimensionality (∂) of computed feature elements gets reduced from $O(d^3) \rightarrow O(d^2)$. The execution workflow of this Algorithm 2 further checks for the noise elements (NAN) in the feature set and also performs correction of feature attributes by replacing the elements of $\vec{X}$ [0:NAN]←NAN by 0. The process of computing the adjusted and normalized feature matrix is shown with simplified execution steps below.

```
Algorithm-3: Normalization of the [val](m x d) ←$\overrightarrow{Lt}$(d)

Start
    1. Create one structure: Feature Set
    2. For each column of $\vec{L}$
        a. $\vec{L}$(m x n x d) →$\overrightarrow{Lt}$(m,1,d)
        b. X(m,1)=[0]
        c. X[val](m x d) ←$\overrightarrow{Lt}$(d)
        d. Check for NAN if found, replace it by 0, X [0:
           NAN]←NAN
End
```

Algorithm-3 computationally executes the normalization of the feature lexicon from the dictionary attributes, making them suitable to train the proposed model with higher efficiency and optimized computation. Here, the appropriate feature lexicons also improve the learning-based performance towards the accuracy of event detection. In this computing stage, the proposed model also incorporates another cost-effective approach of feature lexicon approximation by defining several clusters in k. Here the system invokes a functional segment $f_{\mathbf{NNLS}}$ (): to be operated on the normalized feature lexicon vector X[val] (m x d). Finally, the computed factorize multivariate lexicon blocks of learning-based features are evaluated by the functional component of the non-negative least square approach, which optimizes the non-negative factorization of the lexicon block matrix. The executable functional segment of $f_{\mathbf{NNLS}}$ () is further discussed in the subsequent section. The computed trainable lexicon vector features are then stored in a structure called $\overrightarrow{\mathrm{Fv}}$*({features}* and it is updated. The raw form of Lexicon features Vector $\vec{X}$ (Fig. 1) before NNLS-based lexicon factorization is visualized as follows:

D. The NNLS Algorithm

The explicit function, sparse-"non-negative factorization" as S-NNF, effectively handles multivariate for the depreciation of massive matrix computation's computational resources. The algorithm for S-NNF takes the matrix of the feature vector $\overrightarrow{Fv}$. The dictionary features lexicon block as a mixed-signal to factorize organized as samples in the column and features in a row with clusters (k).

```
Algorithm-4:The NNLS Algorithm
Input:$\overrightarrow{\boldsymbol{Fv}}$ , k
Output: $(\overrightarrow{\boldsymbol{Fv}})\boldsymbol{res}$
Process:
Start:
$\vec{Y}$← Matrix with random value
        for each iteration
                $\vec{A}$← (Moore-pseudo inverse $\vec{Y}$) x $\overrightarrow{\boldsymbol{Fv}}$
                Update normalized ($\vec{A}$)
                $\vec{Y}$←$f$(Ae,$(\overrightarrow{Fv})o$) // function for solution of least square
                        Initialize, a large value of Xp
                        $(\overrightarrow{Fv})c$←$\vec{A}$ x $\vec{Y}$
                        $(\overrightarrow{Fv})f$←$f((\overrightarrow{Fv})p - (\overrightarrow{Fv})c)$ // function for the Frobenius norm
                        Update$(\overrightarrow{Fv})p$←$(\overrightarrow{Fv})c$
                        $(\overrightarrow{Fv})res$←‖$\overrightarrow{Fv}$ - $(\overrightarrow{Fv})c$ ‖
        end
End
```

The NNLS(): The $\vec{X}$ as $\overrightarrow{Fv}$contains -ve elements; therefore, directly non-negative factorization(NNF) is not applicable as in NNF, all $\vec{X}$ and its factors $\vec{W}$ and $\vec{H}$ Shall be having non-negative elements. Therefore, Semi-NNF is used. The semi-non-negative matrix factorization is a technique that learns a low-dimensional dataset representation that lends itself to a clustering interpretation.). A vector $\vec{Y}$as a matrix to hold the coefficient initially consist of random values with the number of rows is equal to the number of clusters(k) and the number of the columns as the number of columns of $\overrightarrow{Fv}$ and another vector $\vec{A}$ As a base matrix. In each iteration of computation, the Moore-pseudo inverse of the Vector $\vec{Y}$ updates the value of $\vec{A}$ after multiplying to the $\overrightarrow{Fv}$ and gets normalized $\vec{A}$. The parameter Ae and $(\overrightarrow{Fv})o$ is taken as an input argument to an algorithm for the solution of the NN-constraint least square problem (LSP) [AR]. With the value of the $\vec{A}$ based on the Ae and $(\overrightarrow{Fv})o$, this algorithm solves for the optimal value of the K in a least-square using equation (4).

$$\mathrm{Ae} = (\overrightarrow{Fv})o \times \mathrm{K} \tag{4}$$

In the problem of Min ‖ Ae - $(\overrightarrow{Fv})o$ x K ‖ such that K≥ 0 for a given Ae and $(\overrightarrow{Fv})o$. A considerable value of the previous fit value $(\overrightarrow{Fv})p$and the current fit value $(\overrightarrow{Fv})c$ = $\vec{A}$ x $\vec{Y}$ provides the Frobenius norm to give the fittest result $(\overrightarrow{Fv})f$ and the current fit value is updated as a previous fit value, and finally, the final residual $(\overrightarrow{Fv})res$ is gets updated using equation (5).

$$(\overrightarrow{Fv})res \leftarrow \| \overrightarrow{Fv} - (\overrightarrow{Fv})c \| \tag{5}$$

The updated value of the final residual as $(\overrightarrow{Fv})res$ at every iteration are computed where the proposed NNLS algorithm optimizes the factorization approach and computes the learning based-features of lexicon blocks in $(\overrightarrow{Fv})res$ after the completion of the training. The proposed system further extracts the test data (GT←$\mathrm{D}_{\mathrm{test}}$) from the repository location (R_1). The process also checks for the GT dimension and converts it into greyscale form with the dimension of (∂). Here the study invokes an explicit function f_{t}(x) to test the S-NNF of the least square, which is obtained as $(\overrightarrow{Fv})res$. The study further discussed the explicit functions of ground truth (GT) computation solution considering the formulated approach.

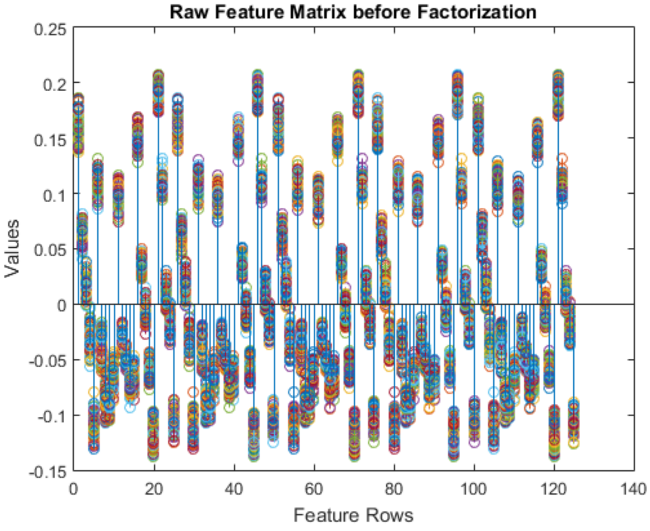


Fig. 1. Visualization of Feature Elements of the Lexicon Vector $\vec{X}$.

Explicit $f_t(x)$ to test the S-NNF of least square with GT

```
Input: GT, $bw$ ←DFL{} , $(\overrightarrow{Fv})res,\ d$
Output: Updated $bw$ ←DFL{}, $d$
Process start:
        Initialize: GT, $bw$ ←DFL{} , $(\overrightarrow{Fv})res,\ d$
        $p \leftarrow R(GT)$
        $C \leftarrow con(1, p)$
        $if\ C \rightarrow$ []
                $Compute: c1 \rightarrow C(row, 1)$
                $Normalize \rightarrow range(\ C1)$
                $Compute: c2 \rightarrow C(row, 2)$
                $Normalize \rightarrow range(\ C2)$
                $update: bw \leftarrow\ Obj(bw)$ $[func\_in: bw, c1, c2]$
        $else$
                $update: bw \leftarrow\ Z[size(GT)]$
        $end$
        $perform\ gt \leftarrow\ bwMO(GT)$
        $bw \leftarrow\ bw \times gt$
        $update: bw \leftarrow DFL(d)$
End
```

In this process, the system initially computes the GT as a binarize form of the D_{test} from the greyscale matrix. The approach here also considers the lexicon properties of DFL for the test sequence generation. Initially, the function computes the region properties from the centroid matrix corresponding to the GT. Further, the function $f_t(x)$ also performs concatenation on the centroid factor of computed region properties and stores the numerical values into a vector *C*. Further algorithmically, the computational process also assesses a conditional check on this computed Vector C. If it finds C as an empty matrix, it computes the 1st column values corresponding to *C* and stores it into another variable, *c1*. The range of *c1* is also get adjusted for ease of computation. Further, the system also computes another variable, c2, from the second column of *C* and adjusts the range of *c2*. Here the function f_t(x) invokes another function to identify the objects in the binary form of an image that overlaps with the pixel attributes (pi). The function here considers bw, c1, and c2 for the execution purpose. Else the process constructs a vector of zeros for bw with the size of the matrix GT. Finally, the computational process applies morphological operations on the pixel of the binary image to identify the Obj and update the bw matrix form. In the ground truth (GT), the contiguous region is considered a connected component, sometimes known as blobs. An example of a label matrix containing the blobs is as in the following Table II:

TABLE II. LABEL MATRIX

1	1	0	2	2	0	3
1	1	0	2	2	0	3

It returns a [1 x n] vector that specifies the center of the region's area. The very first element of the centroid is the x-coordinate and the second element is the y-coordinate. The study further incorporates another functional module, *f_m(x),* to compute the Moore-pseudo inverse $\vec{Y}$ concerning $\overrightarrow{Fv}$ as shown in the above algorithm-4 for NNLS computation and optimization procedure. This explicit function computation is discussed as follows.

Explicit $f_m(x)$ to compute Moore-pseudo inverse $\vec{Y}$

```
Input: $\vec{Y}$ , $\overrightarrow{\boldsymbol{Fv}}$ , $\delta$
Output: $\vec{A}$
Process start:
        Init: $\vec{Y}$ , $\overrightarrow{\boldsymbol{Fv}}$ , $\delta$
        Compute: [$row$, $col$] ←size( $\vec{Y}$) matrix form
        Check: func: input $arg$
        If $arg$ <2
                If ($row$ < $col$)
                        $\vec{A} \leftarrow \frac{(\overrightarrow{\vec{Y}' \times \vec{Y}})}{\vec{Y}'}$
                Else
                        $\vec{A} \leftarrow \frac{\vec{Y}'}{(\overrightarrow{\vec{Y}' \times \vec{Y}})}$
        End
        If ($row$ < $col$)
                $\vec{A} \leftarrow \frac{1}{\delta} \times \sum \mathrm{I(col)}, \vec{Y}'/\vec{Y}' \times \vec{Y}$
        Else
                $\vec{A} \leftarrow \frac{\vec{Y}'}{\frac{1}{\delta} \times \sum \mathrm{I(col)}, \vec{Y}' \times \vec{Y}}$
        End
```

The above function for Moore-pseudo inverse computation is analytically modeled in such a way that it takes $\vec{Y}$, $\overrightarrow{Fv}\ \delta$ as inputs here δ refers to a scalar parameter that produces a stable outcome during the computation, and its value should be tremendous. Further, the process computes the size of the matrix form of $\vec{Y}$ and according to the conditional check, it executes the numerical computation of pseudo-inverse and stores it into $\vec{A}$ in the inverse computed form. The system was further subjected to compute another functional module $f_{norm}(x)$ for the Frobenius norm computation considering $(\overrightarrow{Fv})c$ as input which is the normalized form of the computed Vector $\vec{A}$. The following eq can obtain the computation of the Frobenius norm. (1).

$$\overrightarrow{Fv} \leftarrow \sqrt{\sum\sum(\overrightarrow{Fv})c^2} \qquad (6)$$

The system computes the Frobenius norm during the execution of the proposed NNLS algorithm to strengthen the feature computation and training procedure. It also applies another functional strategy of fast combinatorial to deal with the optimization problem of least square execution mode in NNLS. The function $\boldsymbol{f_O(x)}$ here adjusts the least square problem in $\vec{A}$e and computes a solution matrix k in the form of $\vec{Y}$ During the execution mode of NNLS. For an optimal matrix of $\vec{Y}$ computation, the function $\boldsymbol{f_O(x)}$ takes the input parameters $\vec{A}$e and defined coefficient matrix from $(\overrightarrow{Fv})o$. The optimization problem for the function $\boldsymbol{f_O(x)}$ is formulated as:

$$Min\ f_O(x) \rightarrow ||\vec{A}\mathrm{e}-, (\overrightarrow{Fv})o \times k \qquad (7)$$

$$subjected\ k \geq 0\ for\ given\ \vec{A}\mathrm{e}(\overrightarrow{Fv})o,$$

The proposed solution in $\vec{Y}$ in the form of k obtained from minimizing the function $f_O(x)$, the study here also checked whether the proposed NNLS algorithm converged properly toward the optimality of $\vec{Y}$. The formulated learning model based on NNLS also computes a scalar *K* for linear kernel computation. It constructs a function $f_K(x)$ to compute K, that is, kernel vector concerning column vectors of $c1, c2$. The function $f_K(x)$ is explicitly designed as follows:

Explicit $f_K(x)$ to compute linear kernel matrix

```
Input: $c1, c2$
Output: $K$
Process start:
        Init: $c1, c2$
        If size (c1) >1
                $c1 \leftarrow m(c1,3)$
                $c2 \leftarrow m(c2,3)$
                $c1 \leftarrow c1'$
                $c2 \leftarrow c2'$
        End
        $K \leftarrow c1' \times c2$
End
```

The function here computes a linear kernel matrix to ease the identification of unusual objects during the tracking phase. The formulated approach computes two distinct column vectors and performs normalization of column vectors with a function $m(x)$. Finally, the product of $c1'c2$ are stored into K as a linear kernel matrix. The study constructs another function, $f_{KM}(x)$, to compute the kernel matrix for different kernel functions. Finally, for the computation of the kernel function, the $f_{KM}(x)$ evaluates the kernel function to retain the value of K. The next segment of the study discusses the experimental outcome obtained for different random test instances during the model evaluation and validation phase considering test sequence exploration concerning GT and the context of tracking unusual events.

V. Result Analysis

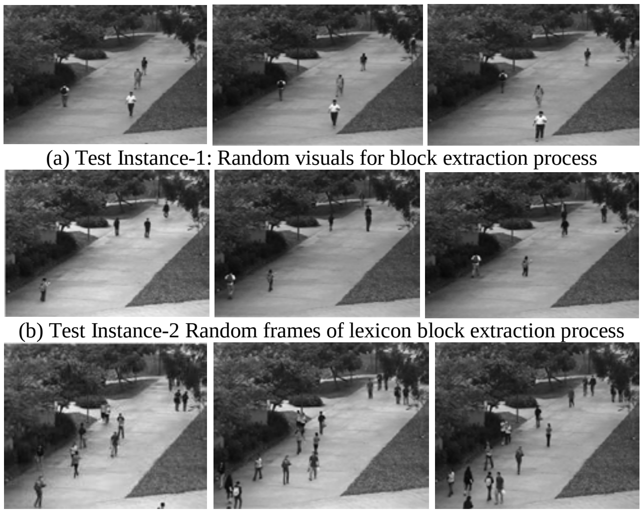
The study performs an extensive numerical analysis by investigating the outcome of a block-shift feature algorithm-based learning model for tracking unusual object movement from a video sequence. Every functional module associated with the formulated approach is evaluated with numerical modeling and a systematic execution flow. The optimized version of the proposed NNLS algorithm for S-NNF converges towards a fixed point. It helps in efficient and faster tracking of the motion patterns associated with the non-pedestrian entities. The study refers to the dataset in [38] for the block-shift feature-based framework's entire design and numerical analysis phase.

A. Analysis of the Dataset for the Experiment

The dataset contains a set of training videos and testing videos as GT for validation, and also it consists of cumulatively 5000 frame sequences for videos. Here, each video data was captured through a camera installed on the roadside to record pedestrians' feeds and the non-pedestrian pattern of movement, which is also considered an unusual movement in this study. Here, during the numerical modeling and computation of scene sequences and dictionary lexicon block extraction, it is realized that each of the moving scene sequences is composed of a set of people walking on streets and moving in two directions. Among the crowd of people, the prime role of the formulated NNLS-based approach is to track significant unusual events in the form of the non-pedestrian pattern of movement and dynamic movement of the pedestrian. The dataset of video sequences also consists of metadata annotation and a GT set of sequences in test data form. Here the annotation form of metadata indicates binary flags for each video scene. The prime intention here is to objectify the significant events to be tracked. Here unusual event tracking belongs to a cart between pedestrians, a wheelchair rolling in sideways, skaters moving in between the walking way, and a biker). The numerical modeling-based framework assesses different training and test instances to validate the performance of an NNLS-based formulated unique event tracking algorithm. It also assesses the algorithm's capability to differentiate the significant events (i.e., unusual events) from normal circumstances (i.e., pedestrians walking on the road). The visualization of the outcome for the video sequence exploration phase partially has already been shown in Fig. 2 above for training and testing sequences of random video scenes. Fig. 2 shows the unexpected visual outcome for feature block extraction during the extended video sequence exploration phase.

Fig. 2 shows the random visuals of scenes involving pedestrians on the road for a different set of training sequences in the event tracking dataset. The data of random visuals are generated during the execution of algorithm-1 for the lexicon block extraction process and feature selection modeling processes. Every algorithm modeling is analytically simplified so that the formulated learning model based on the NNLS algorithm does not encounter convergence problems during the feature evaluation process during run-time. The dataset computed post feature extraction process subjected the NNLS optimization procedure to converge towards optimal DFL with properly trained classes. During the training procedure, the visuals of unusual events are learned, tracked with the appropriate feature modeling, and reposited and indexed with the composition of a matrix set in the subsequent computation process. The proposed system of NNLS-based computation results in a single form of matrix composition with the outcome of training, making the evaluation and validation process computationally faster during run-time. The visuals of the unusual tracing of objects using the block shift feature algorithm are as follows in Fig. 3.

Fig. 3 shows the tracking visuals of the unusual movement of an object using the proposed NNLS algorithm. The validation phase is carried out concerning the computed GT, generated with the explicit function ft(x). To measure the effectiveness of the proposed tracking algorithm, the performance parameters are compared with a highly cited related study by Fang et al. [39], in which the learning model is designed based on deep learning. The comparative outcome in figure xx shows that the formulated approach attains better unusual event tracking accuracy concerning the performance parameters such as recall factor, precision score, specificity score, F1-score, and algorithm execution time. The interpretation of figure xx clearly shows that the formulated NNLS attain a significantly lesser processing time for execution, approximately 0.266621 sec.

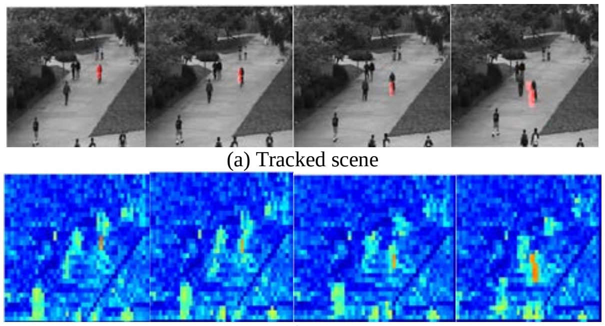


(a) Test Instance-1: Random visuals for block extraction process

(b) Test Instance-2 Random frames of lexicon block extraction process

(c) Test Instance-3: Random visuals of feature lexicon block extraction process

Fig. 2. Random Visuals of the Video Scenes from the Feature Engineering Feature Lexicon Block Extraction Process.



(a) Tracked scene

(b) Color map of the tracked scene

Fig. 3. Tracking Visuals for Training Data-1 in different Frame Instances with the Colormap.

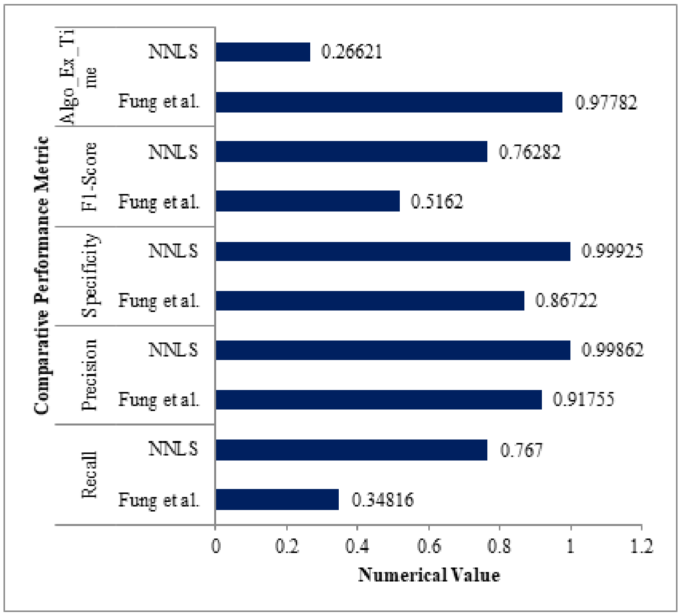


Fig. 4. Comparable Outcomes for the Effectiveness Measure in Comparison with Fang et al. [39].

In contrast, the deep learning-based model in Fang et al. [39] takes a comparatively higher computing time execution of approximately 0.97782 sec. The numerical assessment for the comparative analysis is performed in a similar test environment. The relative performance outcome is shown in the following Fig. 4.

The prime reason behind the effectiveness of the outcome corresponding to the formulated approach NNLS is that it has a lower dependency on the computational resources, unlike the deep learning models. The deep learning-based models have a more considerable dependency on the quantity of the training data set, which can attain accuracy but compromise the computational performance. However, the formulated NNLS-based tracking approach intelligently extracts features and poses lower dependency on the training images with faster execution in run time, making it suitable for real-time video tracking applications.

VI. Conclusion

This paper introduces an efficient scheme of unusual movement tracking from video sequences considering a novel block-shift feature and NNLS algorithm. The study finds a standard dataset of different video sets where the scenes comprise crowd and pedestrian movements. The underlying motive behind this research study is to track the unusual dynamics associated with a mobile object that differs from the pedestrian movement pattern. The proposed research introduces an efficient feature extraction algorithm with a block-wise feature lexicon. It optimizes the NNLS algorithm to optimize the computing and training performance of the learning model. The extensive performance outcome shows that the formulated NNLS-based approach involves training for the learning model, which doesn't include many dependencies on the computational resources and the video feeds, unlike other traditional learning models. Most conventional training models demand a comparatively larger size of trained data to accomplish better tracking accuracy. The NNLS design is optimized so that it can efficiently train with even low or medium training data and performs with higher accuracy in identifying unusual events. The formulated approach's faster and timelier execution makes it highly applicable in a practical environment.

References

[1] R. Fan, F-L Zhang, M.Zhang, "Robust tracking-by-detection using a selection and completion mechanism," Springer, Computational Visual Media, Vol. 3, No. 3, September 2017, 285–294, DOI 10.1007/s41095-017-0083-7.

[2] S. A. Ahmed, D. P. Dogra, S. Kar, and P. P. Roy, "Trajectory-Based Surveillance Analysis: A Survey," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 7, pp. 1985-1997, July 2019, doi: 10.1109/TCSVT.2018.2857489.

[3] Y. Yuan, Y. Lu, and Q. Wang, "Tracking as a Whole: Multi-Target Tracking by Modeling Group Behavior With Sequential Detection," in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 12, pp. 3339-3349, Dec. 2017, doi: 10.1109/TITS.2017.2686871.

[4] C. Liu, H. Chen, K. Lo, C. Wang and J. Chuang, "Accelerating Vanishing Point-Based Line Sampling Scheme for Real-Time People Localization," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 3, pp. 409-420, March 2017, doi: 10.1109/TCSVT.2017.2649019.

[5] D. De Gregorio, R. Zanella, G. Palli, S. Pirozzi and C. Melchiorri, "Integration of Robotic Vision and Tactile Sensing for Wire-Terminal Insertion Tasks," in IEEE Transactions on Automation Science and Engineering, vol. 16, no. 2, pp. 585-598, April 2019, doi: 10.1109/TASE.2018.2847222.

[6] C. Cuevas, R. Martínez, D. Berjón, and N. García, "Detection of Stationary Foreground Objects Using Multiple Nonparametric Background-Foreground Models on a Finite State Machine," in IEEE Transactions on Image Processing, vol. 26, no. 3, pp. 1127-1142, March 2017, doi: 10.1109/TIP.2016.2642779.

[7] T. Huynh-The, C. Hua, N. A. Tu and D. Kim, "Locally Statistical Dual-Mode Background Subtraction Approach," in IEEE Access, vol. 7, pp. 9769-9782, 2019, doi: 10.1109/ACCESS.2019.2891084.

[8] C. Liang and C. Juang, "Moving Object Classification Using a Combination of Static Appearance Features and Spatial and Temporal Entropy Values of Optical Flows," in IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 6, pp. 3453-3464, Dec. 2015, doi: 10.1109/TITS.2015.2459917.

[9] H. Dou, D. Ming, Z. Yang, Z. Pan, Y. Li, and J. Tian, "Object-Based Visual Saliency via Laplacian Regularized Kernel Regression," in IEEE Transactions on Multimedia, vol. 19, no. 8, pp. 1718-1729, Aug. 2017, doi: 10.1109/TMM.2017.2689327.

[10] D. P. Dogra, A. K. Majumdar, S. Sural, J. Mukherjee, S. Mukherjee, and A. Singh, "Toward Automating Hammersmith Pulled-To-Sit Examination of Infants Using Feature Point-Based Video Object Tracking," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 20, no. 1, pp. 38-47, Jan. 2012, doi: 10.1109/TNSRE.2011.2172223.

[11] J. Dunik, O. Straka, M. Simandl, and E. Blasch, "Random-point-based filters: analysis and comparison in target tracking," in IEEE Transactions on Aerospace and Electronic Systems, vol. 51, no. 2, pp. 1403-1421, April 2015, doi: 10.1109/TAES.2014.130136.

[12] M. Gupta, S. Kumar, L. Behera, and V. K. Subramanian, "A Novel Vision-Based Tracking Algorithm for a Human-Following Mobile Robot," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, no. 7, pp. 1415-1427, July 2017, doi: 10.1109/TSMC.2016.2616343.

[13] T. Zhang and S. Fei, "Improved particle filter for object tracking," 2011 Chinese Control and Decision Conference (CCDC), Mianyang, 2011, pp. 3586-3590, doi: 10.1109/CCDC.2011.5968843.

[14] Jong-Min Jeong, Tae-Sung Yoon, and Jin-Bae Park, "Kalman filter-based multiple objects detection-tracking algorithm robust to occlusion," 2014 Proceedings of the SICE Annual Conference (SICE), Sapporo, 2014, pp. 941-946, doi: 10.1109/SICE.2014.6935235.

[15] N. Widynski and M. Mignotte, "A MultiScale Particle Filter Framework for Contour Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 10, pp. 1922-1935, Oct. 2014, doi: 10.1109/TPAMI.2014.2307856.

[16] Shen, Chunhua& Kim, Junae& Wang, Hanzi. (2010). Generalized Kernel-Based Visual Tracking. Circuits and Systems for Video Technology, IEEE Transactions on. 20. 119 - 130. 10.1109/TCSVT.2009.2031393.

[17] Chen, Zezhi&Husz, Zsolt& Wallace, Iain & Wallace, Andrew. (2007). Video Object Tracking Based on a Chamfer Distance Transform. Proceedings - International Conference on Image Processing, ICIP. 3. 357-360. 10.1109/ICIP.2007.4379320.

[18] Y. Wang and J. Zhang, "Application of SVM in Object Tracking Based on Laplacian Kernel Function," 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, 2016, pp. 557-561, doi: 10.1109/IHMSC.2016.121.

[19] K. Sun, W. Tao and Y. Qian, "Guide to Match: Multi-Layer Feature Matching With a Hybrid Gaussian Mixture Model," in IEEE Transactions on Multimedia, vol. 22, no. 9, pp. 2246-2261, Sept. 2020, doi: 10.1109/TMM.2019.2957984.

[20] Q. Lin et al., "Robust Stereo-Match Algorithm for Infrared Markers in Image-Guided Optical Tracking System," in IEEE Access, vol. 6, pp. 52421-52433, 2018, doi: 10.1109/ACCESS.2018.2869433.

[21] Q. Zhu, H. Xiong, and X. Jiang, "Shape-oriented segmentation with graph matching corroboration for silhouette tracking," 2012 Visual Communications and Image Processing, San Diego, CA, 2012, pp. 1-6, doi: 10.1109/VCIP.2012.6410762.

[22] Z. Boukhers, K. Shirahama, and M. Grzegorzek, "Example-Based 3D Trajectory Extraction of Objects From 2D Videos," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 9, pp. 2246-2260, Sept. 2018, doi: 10.1109/TCSVT.2017.2727963.

[23] S. Damotharasamy, "Approach to model human appearance based on sparse representation for human tracking in surveillance," in IET Image Processing, vol. 14, no. 11, pp. 2383-2394, 18 9 2020, doi: 10.1049/iet-ipr.2018.5961.

[24] L. Kong, D. Huang, J. Qin, and Y. Wang, "A Joint Framework for Athlete Tracking and Action Recognition in Sports Videos," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 2, pp. 532-548, Feb. 2020, doi: 10.1109/TCSVT.2019.2893318.

[25] Y. Lee, Z. Tang and J. Hwang, "Online-Learning-Based Human Tracking Across Non-Overlapping Cameras," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 10, pp. 2870-2883, Oct. 2018, doi: 10.1109/TCSVT.2017.2707399.

[26] C. Liu, R. Yao, S. H. Rezatofighi, I. Reid and Q. Shi, "Model-Free Tracker for Multiple Objects Using Joint Appearance and Motion Inference," in IEEE Transactions on Image Processing, vol. 29, pp. 277-288, 2020, doi: 10.1109/TIP.2019.2928123.

[27] A. Ratre and V. Pankajakshan, "Tucker tensor decomposition-based tracking and Gaussian mixture model for anomaly localization and detection in surveillance videos," in IET Computer Vision, vol. 12, no. 6, pp. 933-940, 9 2018, doi: 10.1049/iet-cvi.2017.0469.

[28] S. Tian, X. Yin, Y. Su, and H. Hao, "A Unified Framework for Tracking Based Text Detection and Recognition from Web Videos," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 3, pp. 542-554, 1 March 2018, doi: 10.1109/TPAMI.2017.2692763.

[29] C. Yang et al., "Tracking Based Multi-Orientation Scene Text Detection: A Unified Framework With Dynamic Programming," in IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3235-3248, July 2017, doi: 10.1109/TIP.2017.2695104.

[30] S. C. Wong, V. Stamatescu, A. Gatt, D. Kearney, I. Lee, and M. D. McDonnell, "Track Everything: Limiting Prior Knowledge in Online Multi-Object Recognition," in IEEE Transactions on Image Processing, vol. 26, no. 10, pp. 4669-4683, Oct. 2017, doi: 10.1109/TIP.2017.2696744.

[31] X. Cheng, Y. Gu, B. Chen, Y. Zhang, and J. Shi, "Weighted Multiple Instance-Based Deep Correlation Filter for Video Tracking Processing," in IEEE Access, vol. 7, pp. 161220-161230, 2019, doi: 10.1109/ACCESS.2019.2951600.

[32] A. Khan et al., "Forensic Video Analysis: Passive Tracking System for Automated Person of Interest (POI) Localization," in IEEE Access, vol. 6, pp. 43392-43403, 2018, doi: 10.1109/ACCESS.2018.2856936.

[33] X. Lan, M. Ye, R. Shao, B. Zhong, P. C. Yuen, and H. Zhou, "Learning Modality-Consistency Feature Templates: A Robust RGB-Infrared Tracking System," in IEEE Transactions on Industrial Electronics, vol. 66, no. 12, pp. 9887-9897, Dec. 2019, doi: 10.1109/TIE.2019.2898618.

[34] X. Liu, Y. Xu, L. Zhu, and Y. Mu, "A Stochastic Attribute Grammar for Robust Cross-View Human Tracking," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 10, pp. 2884-2895, Oct. 2018, doi: 10.1109/TCSVT.2017.2781738.

[35] Van Benthem MH, Keenan MR. A fast algorithm for the solution of large-scale non-negativity-constrained least squares problems. Journal of Chemometrics: A Journal of the Chemometrics Society. 2004 Oct;18(10):441-50.

[36] Karanam Sunil Kumar and N P Kavya, "Compact Scrutiny of Current Video Tracking System and its Associated Standard Approaches" International Journal of Advanced Computer Science and Applications (IJACSA), 11(12) 2020. http://dx.doi.org/10.14569/IJACSA.2020.0111 249.

[37] Kumar, K.S. and Kavya, N.P., 2021, April. Novel Approach of Video Tracking System Using Learning-Based Mechanism over Crowded Environment. In Computer Science On-line Conference (pp. 67-76). Springer, Cham.

[38] http://www.svcl.ucsd.edu/projects/anomaly/dataset.html.

[39] Z. Fang & F. Fei, & Y. Fang, & C. Lee, & N. Xiong, & L. Shu, & S. Chen, "Abnormal event detection in crowded scenes based on deep learning", Springer Journal of Multimedia Tool Application, 2016, DOI 10.1007/s11042-016-3316-3.

An Ontological Model based on Machine Learning for Predicting Breast Cancer

Hakim El Massari[1]*, Noreddine Gherabi[2], Sajida Mhammedi[3], Hamza Ghandi[4], Fatima Qanouni[5], Mohamed Bahaj[6]
National School of Applied Sciences, Sultan Moulay Slimane University, Lasti Laboratory, Khouribga, Morocco[1, 2, 3, 4, 5]
Faculty of Sciences and Technologies, Hassan First University, Settat, Morocco[6]

***Abstract*—Breast cancer is mostly a female disease, but it may affect men as well even at a considerably lower percentage. An automated diagnosis system should be built for early detection because manual breast cancer diagnosis takes a long time. Doctors have lately achieved significant advances in the early identification and treatment of breast cancer in order to decrease the rate of mortality caused by the latter. Researchers, on the other hand, are analysing large amounts of complicated medical data by employing a combination of statistical and machine learning methodologies to assist clinicians in predicting breast cancer. Various machine learning approaches, including ontology-based Machine Learning methods, have lately played an essential role in medical science by building an automated system that can identify breast cancer. This study examines and evaluates the most popular machine learning algorithms, besides the ontological model based on Machine Learning. Among the classification methods investigated were Naive Bayes, Decision Tree, Logistic Regression, Support Vector Machine, Artificial Neural Network, Random Forest, and k-Nearest Neighbours. The dataset utilized has 683 instances and is available for download from the Kaggle website. The findings are assessed using performance measures generated from the confusion matrix, such as F-Measure, Accuracy, Precision, and Recall. The ontology model surpassed all machine learning techniques, according to the results.**

Keywords—Machine learning; prediction; ontology; semantic web rule language; decision tree; breast cancer

I. Introduction

In 2020, 2.3M women were identified with breast cancer, with 685 thousand fatalities worldwide. By the end of 2020, there will have been 7.8M women diagnosed with breast cancer in the previous five years, making it the most frequent kind of cancer in the world. Breast cancer claims more DALYs from women than any other kind of cancer worldwide. Breast cancer affects women of all ages after puberty in every country throughout the world, but at a rising rate in the latter stages of life. From the 1930s through the 1970s, there was minimal change in the breast cancer death rate. In nations where early diagnosis systems are available in conjunction with various treatment options to remove intrusive illnesses, life expectancies started to get better as in 1980s.

Machine learning (ML) is one of the most constantly evolving areas of computer science, with a wide range of applications [1], [2]. It is the process of obtaining usable information from a big quantity of data [3]. Marketing, Industry, Medical diagnosis, and other scientific domains all make use of ML approaches. ML algorithms are well-suited for medical data analysis since they have been frequently employed in medical datasets. ML comes in several forms, including classification, regression, and clustering. Each form has a particular consequence and influence depending on the problem that we are attempting to address. We focus on classification algorithms in our work because of their high accuracy and performance in classifying a given dataset into predetermined categories and predicting future events or information from that data. In the medical field, classification algorithms are often utilized, particularly in the diagnosis of illnesses such as breast cancer. As a result, regularly used machine learning classification methods such as Support Vector Machine (SVM), k-Nearest Neighbours (KNN), Artificial Neural Network (ANN), Naive Bayes (NB), Logistic regression (LR), and Decision Tree (DT) are utilized to detect patients with breast cancer at an early stage.

Several researches have been conducted, and various machine learning models have been implemented, to identify and predict breast cancer diagnoses [4]. For example, this study [5] sought to identify the most accurate machine learning approaches for predicting patients with breast cancer. Various supervised machine-learning algorithms, including RF, KNN, DT, AdaboostM1, LR, and ANN [6], were used and their performance was compared. The authors of this study [7], examined linear discriminant analysis with support vector machines in terms of specificity, F-Measure, accuracy, and sensitivity to see which method is better for classifying breast cancer datasets. The results reveal that the support vector machine outperforms the linear discriminant analysis, with a 98.77 percent accuracy rate. [8] This study covers the whole Bayesian technique to assess the predicted distribution of all classes using three datasets and three classifiers: naive bayes (NB), bayesian networks (BN), and tree augmented naive bayes (TAN). The outcomes showed that the BN method performed the best, with an accuracy rate of 97 percent.

Breast cancer diagnosis and prediction has garnered a lot of attention in recent years, and numerous ways have been taken to address this issue [9]–[12]. The present focus is on machine learning and the semantic web. The Wisconsin Breast Cancer Dataset was used in this study [13]. The authors' purpose is to study the dataset and assess the efficacy of several ML algorithms for predicting breast cancer. Several machine learning methods have been developed to differentiate benign and malignant tumors. To predict whether breast cancer is malignant or benign, [14] the authors used machine learning and computer vision to extract features and construct an optimized model by varying hyper-parameter values. The

*Corresponding Author.

analysis is carried out using a support vector machine, and several quality indicators are produced, with the observed findings being notable. [15] The goal of this study is to evaluate the classification algorithms' prediction accuracy in terms of efficiency and effectiveness. The authors conduct a rigorous comparison of classification algorithms such as SVM, DT, NB, and RF in terms of prediction accuracy utilizing WEKA and a 10-fold cross validation approach on the Wisconsin Diagnostic Breast Cancer dataset. In this study [16], two machine learning algorithms, Decision Tree Classifier, and Logistic Regression, were implemented for breast cancer prediction on the "Breast Cancer Wisconsin (Diagnostic) Data Set," and their accuracies were compared and the Decision Tree Classifier is the most suited-algorithm for prediction since it has a precise prediction accuracy.

Recently, researchers published a significant quantity of research utilizing machine-learning algorithms to diagnose breast cancer [17]–[20]. In this comparison study [21], 4 machine learning (ML) algorithms were used: DT classifiers, SVM, KNN, and RF, and the results show that Support Vector Machine has the highest accuracy of 97% among them for the classification of breast tumors in women. On the Wisconsin Breast Cancer dataset, [22] the authors examined five supervised machine learning algorithms: LR, RF, KNN, ANN, and SVM. The metrics of the confusion matrix are used to assess the study's performance. According to the results, the ANNs had the greatest accuracy score of 98.57 percent.

Furthermore, ontology has been one of the most widely used techniques to managing, organizing, and extracting data throughout the last few decades. It is a way of data representation that has been effectively utilized in a number of domains, particularly the medical domain. It is significant in computer science because of its ability to express many concepts and their relationships across fields. In reality, no single ontology is sufficient to meet today's expanding healthcare demands, and ontologies must be combined with machine learning algorithms to facilitate data integration and analysis. The authors in [23] created and explored an ontology-based decision tree model able to predict diabetes, [24] then compared the findings to numerous ML techniques, and discovered that the ontology model outperforms all other classifiers.

In this research, we intend to compare seven prominent classification approaches with the Ontological Model using carefully chosen criteria obtained from the confusion matrix, such as F-Measure, Accuracy, Precision, and Recall. The rest of this paper is organized as follows: Section II describes the methodologies utilized in this comparison analysis. Section III summarizes the findings and discussion. Section IV concludes and discusses future work.

II. Methods and Evaluation

The approaches and materials employed, as well as the experimental methodology, dataset description, machine learning algorithms, ontology model, and evaluation metrics, are all included in this section. Fig. 1 depicts the process flowchart for this comparative study.

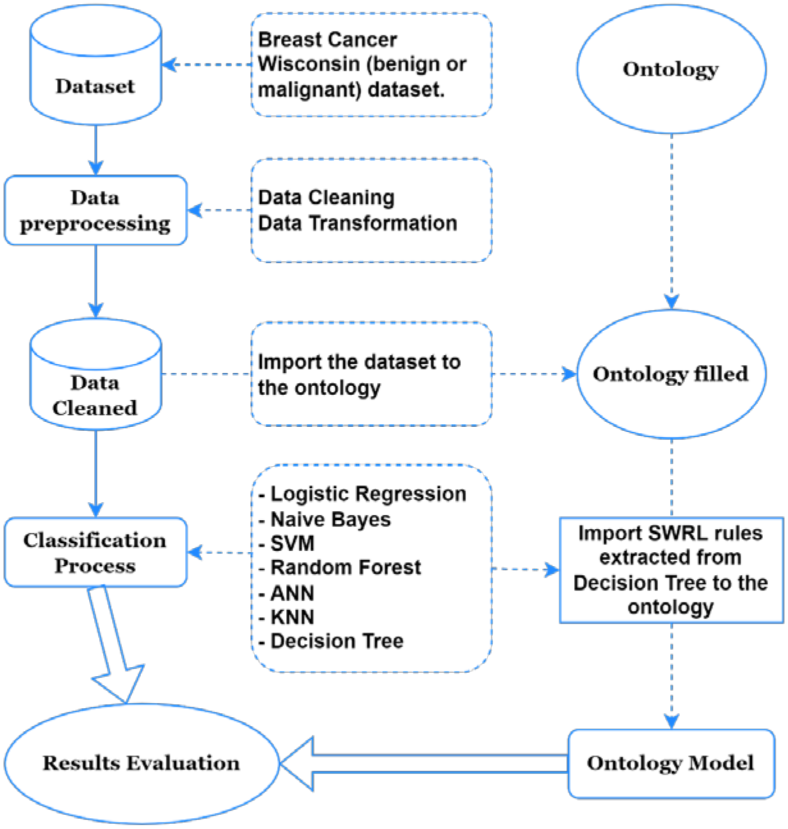


Fig. 1. Experimental Workflow.

A. *Data Preprocessing*

The dataset used is Breast Cancer Wisconsin - benign or malignant from Kaggle website, it consists of 683 instances and 10 features (9 attributes and the last one is a target). A full description of all dataset attributes is provided in Table I.

To build an effective machine learning classifier, we should always start with data cleaning, normalization of features, transformation of features, and even creation of new features from the dataset. The dataset contains 234 similar instances, after removing duplicated instances the remaining is 449 instances, where 213 represent benign cancer cells and 236 represent malignant cancer cells. We would like to inform you that in order to provide a fair comparison of the classification results obtained, we did not use any feature selection or performance-boosting methods.

B. *Machine Learning Algorithms*

We have used Weka software for all machine learning algorithms to predict whether the cancer cells are benign or malignant. Weka comprises tools for data classification, clustering, visualization, preparation, association rules mining, and regression [25].

We used the seven most classifiers used to classify datasets (Decision Tree, Random Forest, Logistic Regression, Artificial Neural Network, Naïve Bayes, Support Vector Machine, k-Nearest Neighbours). In addition, we employed two modes of test options: 10-fold cross validation and percentage split (split 50% train, remainder test) for the reason of enriching the study.

C. *Ontological Model*

This section presents the technologies used to create the ontology, besides the approach used to build the ontology model with the help of rules extracted from DT. This methodology has been referred to in this research for more details [26], which we recommend reading for more information. We'll go through some specifics shortly here.

1) Ontology construction: The ontology was built using the Protégé software, which is an open-source platform that provides a set of tools to a growing user community for constructing domain models and knowledge-based applications with ontologies. The ontology was created manually; the main classes are Diagnostic and Patient. The graphical representation of the ontology is shown in Fig. 2.

TABLE I. DATASET FEATURE'S INFORMATION

Attribute	**Description**
1- clump	Clump Thickness: Benign cells often form monolayers, whereas malignant cells frequently form multilayers.
2- ucz	Uniformity of Cell Size: Cancer cells differ in size.
3- ucp	Uniformity of Cell Shape: Cancer cells differ in form.
4- adhesion	Marginal Adhesion: Adhesion loss is an indication of cancer.
5- epithelial	Single Epithelial Cell Size: Is connected to the previously mentioned uniformity. Significantly expanded epithelial cells may be cancerous cells
6- bare_nuclei	Bare Nuclei: These are common in benign tumors.
7- bland_chromatin	Bland Chromatin: In benign cells, the nucleus has a homogenous texture.
8- normal_nucleoli	Normal Nucleoli: In normal cells, the nucleolus is generally quite tiny, if at all detectable. The nucleoli grow more visible in cancer cells.
9- mitoses	-
10- Class	Predicted class (2 for benign, 4 for malignant).

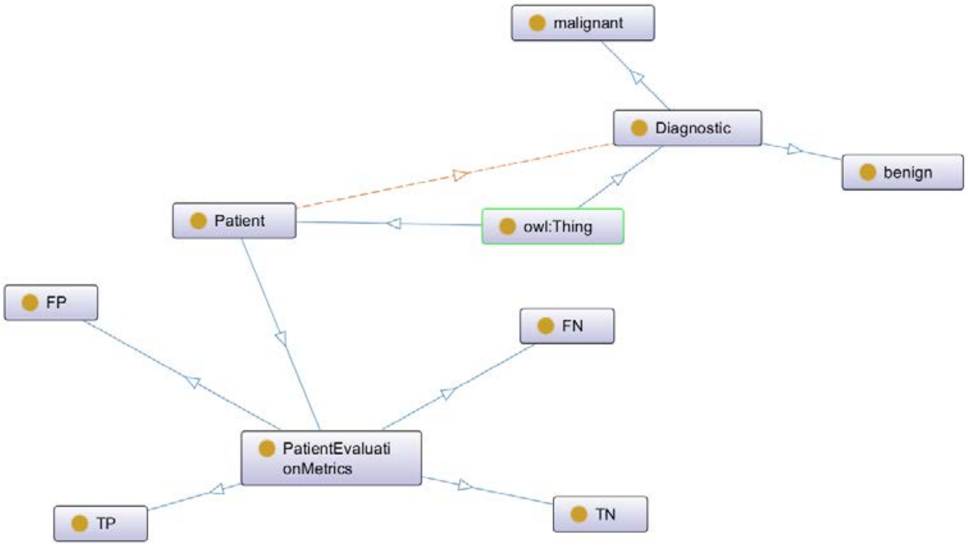
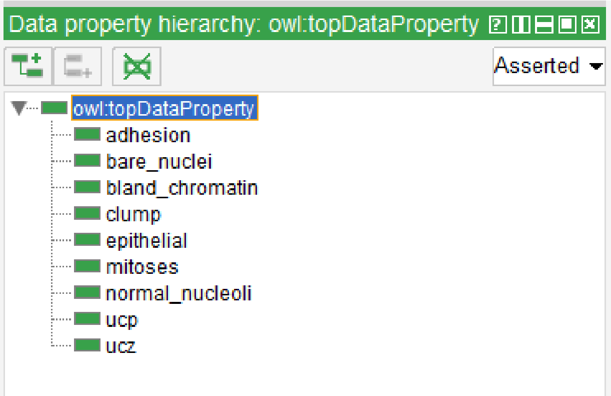


Fig. 2. The Ontology Graph.



Data property hierarchy: owl:topDataProperty

- owl:topDataProperty
  - adhesion
  - bare_nuclei
  - bland_chromatin
  - clump
  - epithelial
  - mitoses
  - normal_nucleoli
  - ucp
  - ucz

Fig. 3. Data Properties.

2) Data properties and instances: The data properties used in the ontology are the same attributes presented in Table I which are used to build models of machine learning algorithms. Fig. 3 illustrates the data properties. A plugin among the Protégé software plugins called Cellfie is used to import the same dataset used in Weka.

3) Semantic web language rules and pellet reasoned: Following the creation of classes, data properties, and instances in the ontology. We need to establish the SWRL reasoning rules. To achieve this, we used the SWRLTab plugin, we retrieved created rules from the DT algorithm, and imported them into Protégé. The collected rules from the DT algorithm are converted using the Java programming language, with each leaf of the tree extracted as a single SWRL rule. For instance/

<u>A leaf from the DT algorithm</u>
If ucp > 2 && ucz ≤ 4 && bare_nuclei ≤ 2 && adhesion ≤ 3 THEN put the patient in benign

<u>SWRL resulted</u>
Patient(?P) ^ ucp(?P, ?UCP) ^ swrlb:greaterThan(?UCP, '2'^^xsd:decimal) ^ ucz(?P, ?UCZ) ^ swrlb:lessThanOrEqual(?UCZ, '4'^^xsd:decimal) ^ bare_nuclei(?P, ?BN) ^ swrlb:lessThanOrEqual(?BN, '2'^^xsd:decimal) ^ adhesion(?P, ?A) ^ swrlb:lessThanOrEqual(?A, '3'^^xsd:decimal) → benign

To execute SWRL rules and infer new ontology axioms we utilized another plugin from Protégé software named Pellet [27], which includes capabilities for checking ontology coherence, deals with SWRL rules, computing the classification hierarchy, deals with OWL, explaining inferences, and answering SPARQL queries. It uses the Ontology and SWRL rules to initiate the inference and then determines if the cancer cells are benign or malignant. The ontology classifier's results are reported in the next section.

D. Evaluation

ROC Area, F-Measure, Root mean squared error, Recall, Accuracy, Root relative squared error, Precision, Kappa statistic, and other performance measures are employed to assess ML algorithms. We employed two test modes (split-test and K-fold cross-validation) using several metrics including Recall, F-Measure, Accuracy, and Precision to analyze our experimental results, which are presented below and in Fig. 4. Furthermore, the same criteria are utilized to assess the validity of this comparison research including ML classifiers and the ontological model.

$$\mathrm{ACC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (1)$$

$$\mathrm{PREC} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (2)$$

$$REC = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F-Measure} = 2 * \frac{PREC * REC}{PREC+REC} \quad (4)$$

Other metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), are available but are most commonly employed in regression issues. As a result, owing to classification issues imposed by the dataset and techniques employed.

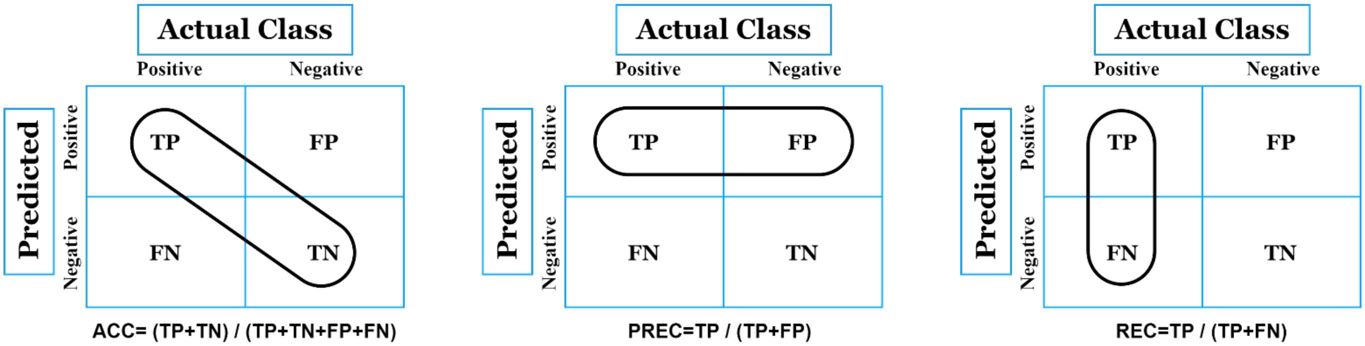


Fig. 4. Performance Metrics.

III. Results and Discussion

In this section, the results of the evaluation of the various classifiers that were used in this study are presented. The statistics and results of the ontological model are also shown in Tables II, III, and Fig. 5 illustrates the performance metrics of the ontology model.

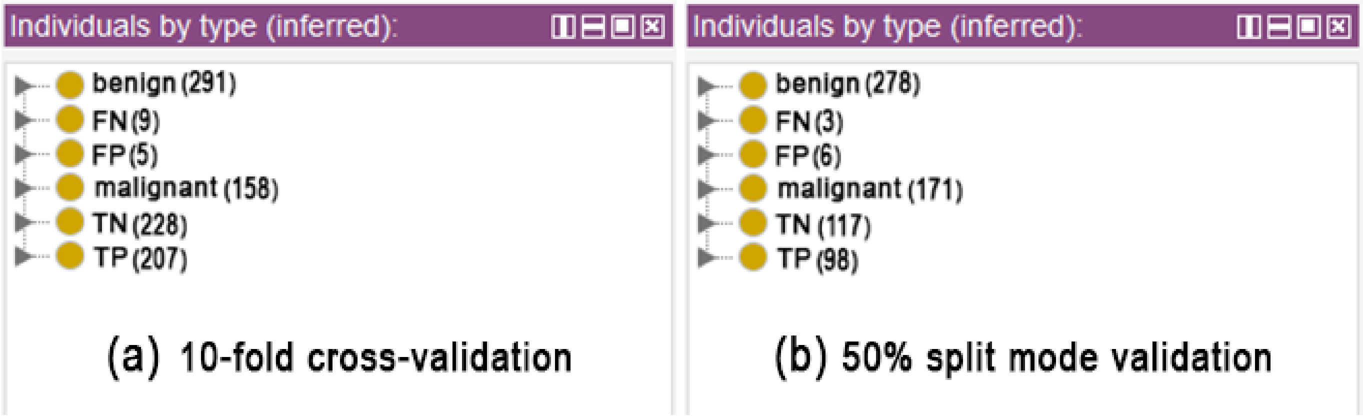
The results of this study provide a visual representation of the various metrics that are used in this research, such as precision, F-measure, Recall, and Accuracy, as shown in Fig. 6-9. Table IV also shows the results of the various classifiers that were used in this research.

TABLE II. 10-Fold Cross-Validation for Ontological Model

Confusion matrix		Actual class	
		positive	Negative
Predicted class	positive negative	TP : 207 FN : 9	FP : 5 TN : 228

TABLE III. 50% Split Mode for Ontological Model

Confusion matrix		Actual class	
		positive	negative
Predicted class	positive negative	TP : 98 FN : 3	FP : 6 TN : 117



Individuals by type (inferred):

- benign (291)
- FN (9)
- FP (5)
- malignant (158)
- TN (228)
- TP (207)

(a) 10-fold cross-validation

Individuals by type (inferred):

- benign (278)
- FN (3)
- FP (6)
- malignant (171)
- TN (117)
- TP (98)

(b) 50% split mode validation

Fig. 5. Results of Inferred Concepts.

Accuracy:

The ontological model achieved the maximum value of 96.88% and Random Forest with rate of 96.00%, and 95.30 % for both Support Vector Machine and Logistic Regression in terms of 10-fold cross-validation, according to Fig. 6 and Table IV.

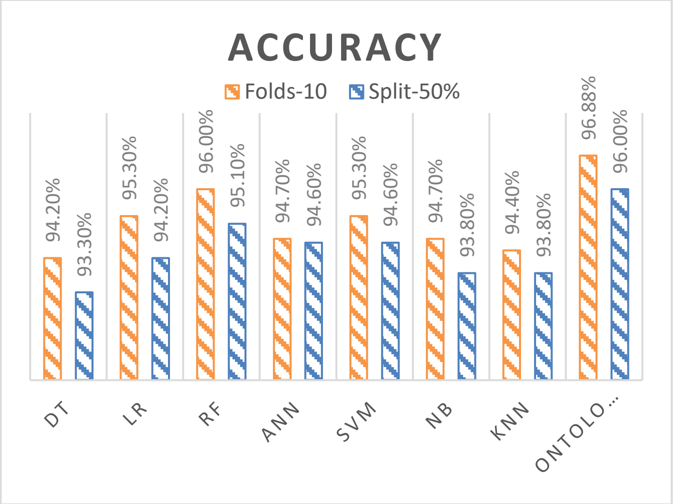


Fig. 6. Comparison Results of Accuracy.

Almost the same results using split test mode, we obtained 96.00%, 95.10% for Ontology and Random Forest consecutively, and 94.60% for both Support Vector Machine and Artificial Neural Network.

Precision:

The ontology classifier has the highest Precision of 97.64% in terms of 10-fold cross-validation mode, followed by Random Forest and Naïve Bayes. Concerning split test mode, the highest Precision value of 97.00% goes for ANN. More details are shown in Table IV and Fig. 7.

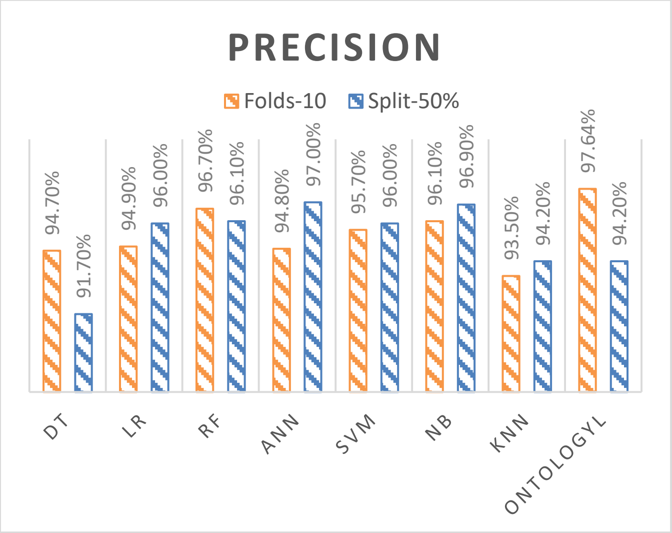


Fig. 7. Comparison Results of Precision.

Recall:

According to Fig. 8 and Table IV, the ontological model has the highest Recall values of 95.83 % and 97.00 % for both test modes, followed by RF, LR, and KNN for 10-fold cross-validation mode, and Decision Tree and Random Forest for split test mode.

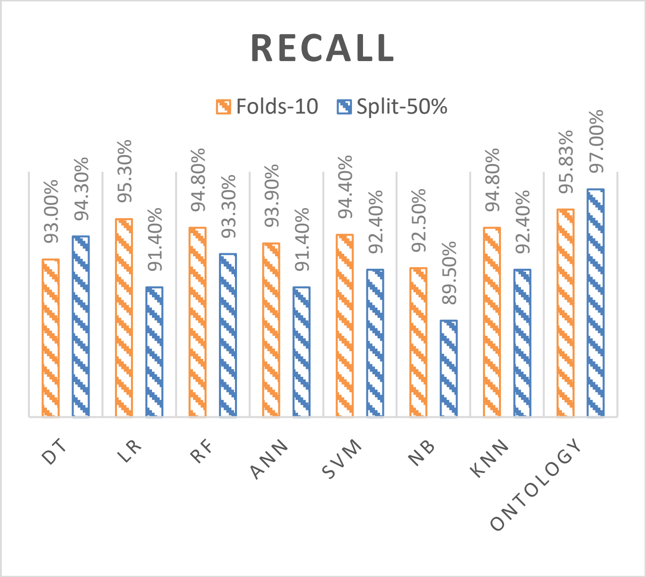


Fig. 8. Comparison Results of Recall.

F-Measure:

According to Fig. 9 and Table IV, the ontology model had the greatest value of 96 % in both test modes, followed by Random Forest in the second position and Support Vector Machine in the third position.

The experimental findings reveal that the ontology model has the highest accuracy of 96.9 %, followed by the Random Forest at 96.00 % and both Logistic Regression and Support Vector Machine at 95.30 %. In terms of the data stated above, we see no significant difference between 50%-Split and 10-Folds test modes. We conclude that the ontological model can aid by extending the scope machine learning model. They can comprise any data kind or variation, and each diver data can be assigned to a certain job. Combining the ontological model with machine learning may provide well outcomes. The ontological model achieves results that are comparable to machine learning classifiers. Humans may interpret the findings, and the rules can be modified or added as needed. Furthermore, it supports unstructured, semi-structured, and structured data formats, allowing for more seamless data integration. It can comprise all aspects of the data modeling process, starting with schemas at the most basic level. As a result, they can handle the massive amounts of data utilized as input for machine learning training or output as outcomes. Furthermore, ontology matches any organization's aim, which might be mathematical, logical, or semantic-based. To the best of our knowledge, this is the first comparative study of the ontological model and ML in which we have integrated the ontology with ML, especially in the area of breast cancer detection. As a result, no significant comparison can be done.

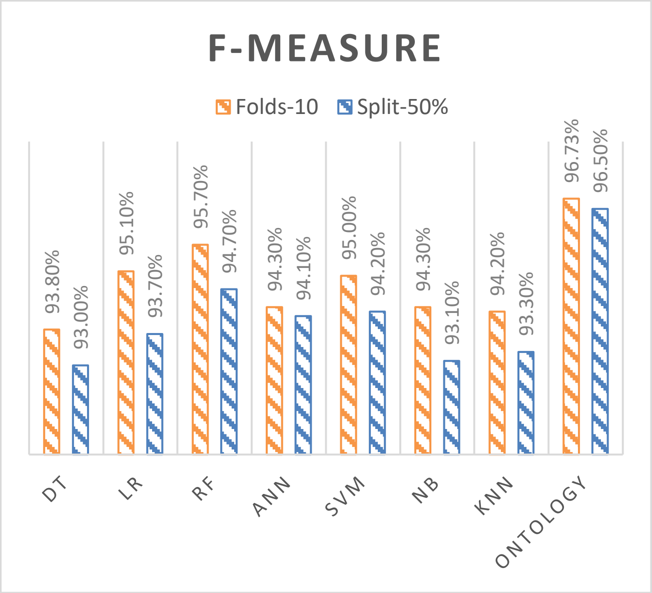


Fig. 9. F-Measure Comparison Findings.

TABLE IV. ONTOLOGICAL MODEL AND MACHINE LEARNING CLASSIFIERS RESULTS

	Accuracy		**Precision**		**Recall**		**F-Measure**	
	Folds-10	Split-50%	Folds-10	Split-50%	Folds-10	Split-50%	Folds-10	Split-50%
DT	0.942	0.933	0.947	0.917	0.93	0.943	0.938	0.93
LR	0.953	0.942	0.949	0.96	0.953	0.914	0.951	0.937
RF	0.96	0.951	0.967	0.961	0.948	0.933	0.957	0.947
ANN	0.947	0.946	0.948	0.97	0.939	0.914	0.943	0.941
SVM	0.953	0.946	0.957	0.96	0.944	0.924	0.95	0.942
NB	0.947	0.938	0.961	0.969	0.925	0.895	0.943	0.931
KNN	0.944	0.938	0.935	0.942	0.948	0.924	0.942	0.933
Ontology	0.969	0.960	0.976	0.942	0.958	0.970	0.967	0.965

IV. Conclusion

ML methods are widely employed in all scientific disciplines and have revolutionized industries all over the world. The use of machine learning techniques and algorithms in healthcare has recently advanced significantly. These approaches have shown success and may be valuable in the treatment of enduring diseases such as breast cancer. Furthermore, the Semantic Web has proven its usefulness and effectiveness in a multitude of areas, including health. As a Semantic Web component, ontology has the capability to treat concepts and relationships in the same way that humans view connected concepts.

In this research, we provided seven machine learning algorithms and an ontology model, as well as a comparison of their performance. Furthermore, two test modes are employed: 10-fold cross validation and percentage split, and several performance measures such as Accuracy, F-Measure, Precision, and Recall are employed to assess the outcomes. The findings show that the ontological model has the uppermost accuracy even when no feature selection is used. This brings us to a new search area, to which we advise and urge academics to participate and produce new insights in the same context, in order to provide additional outcomes and analysis, in order to make a forecast, recommendation, or decision, and so on. In future work, we want to improve this comparison analysis by adopting new ways to incorporate ML rules with the ontological model method, as well as regression machine learning algorithms.

References

[1] L. Rubinger, A. Gazendam, S. Ekhtiari, and M. Bhandari, “Machine learning and artificial intelligence in research and healthcare☆,☆☆,” Injury, Feb. 2022, doi: 10.1016/j.injury.2022.01.046.

[2] H. Abdel-Jaber, D. Devassy, A. Al Salam, L. Hidaytallah, and M. EL-Amir, “A Review of Deep Learning Algorithms and Their Applications in Healthcare,” Algorithms, vol. 15, no. 2, Art. no. 2, Feb. 2022, doi: 10.3390/a15020071.

[3] H. El Massari, S. Mhammedi, N. Gherabi, and M. Nasri, “Virtual OBDA Mechanism Ontop for Answering SPARQL Queries Over Couchbase,” in Advanced Technologies for Humanity, Cham, 2022, pp. 193–205. doi: 10.1007/978-3-030-94188-8_19.

[4] N. Al-Azzam and I. Shatnawi, “Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer,” Ann. Med. Surg., vol. 62, pp. 53–64, Feb. 2021, doi: 10.1016/j.amsu.2020.12.043.

[5] K. Rajendran, M. Jayabalan, and V. Thiruchelvam, “Predicting Breast Cancer via Supervised Machine Learning Methods on Class Imbalanced Data,” Int. J. Adv. Comput. Sci. Appl. IJACSA, vol. 11, no. 8, Art. no. 8, 31 2020, doi: 10.14569/IJACSA.2020.0110808.

[6] M. Alshammari and M. Mezher, “A Comparative Analysis of Data Mining Techniques on Breast Cancer Diagnosis Data using WEKA Toolbox,” Int. J. Adv. Comput. Sci. Appl. IJACSA, vol. 11, no. 8, Art. no. 8, 31 2020, doi: 10.14569/IJACSA.2020.0110829.

[7] Z. Rustam, Y. Amalia, S. Hartini, and G. S. Saragih, “Linear discriminant analysis and support vector machines for classifying breast cancer,” IAES Int. J. Artif. Intell. IJ-AI, vol. 10, no. 1, Art. no. 1, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp253-256.

[8] W. N. L. W. H. Ibeni, M. Z. M. Salikon, A. Mustapha, S. A. Daud, and M. N. M. Salleh, “Comparative analysis on bayesian classification for breast cancer problem,” Bull. Electr. Eng. Inform., vol. 8, no. 4, Art. no. 4, Dec. 2019, doi: 10.11591/eei.v8i4.1628.

[9] F. S. Khan, M. I. Abbasi, M. Khurram, M. N. H. Mohd, and M. D. Khan, “Breast cancer histological images nuclei segmentation and optimized classification with deep learning,” Int. J. Electr. Comput. Eng. IJECE, vol. 12, no. 4, Art. no. 4, Aug. 2022, doi: 10.11591/ijece.v12i4.pp4099-4110.

[10] N. F. Idris, M. A. Ismail, M. S. Mohamad, S. Kasim, Z. Zakaria, and T. Sutikno, “Breast cancer disease classification using fuzzy-ID3 algorithm based on association function,” IAES Int. J. Artif. Intell. IJ-AI, vol. 11, no. 2, Art. no. 2, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp448-461.

[11] T. S. Lim, K. G. Tay, A. Huong, and X. Y. Lim, “Breast cancer diagnosis system using hybrid support vector machine-artificial neural network,” Int. J. Electr. Comput. Eng. IJECE, vol. 11, no. 4, Art. no. 4, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3059-3069.

[12] R. A. I. Alhayali, M. A. Ahmed, Y. M. Mohialden, and A. H. Ali, “Efficient method for breast cancer classification based on ensemble hoffeding tree and naïve Bayes,” Indones. J. Electr. Eng. Comput. Sci., vol. 18, no. 2, Art. no. 2, May 2020, doi: 10.11591/ijeecs.v18.i2.pp1074-1080.

[13] S. Ara, A. Das, and A. Dey, “Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms,” in 2021 International Conference on Artificial Intelligence (ICAI), Apr. 2021, pp. 97–101. doi: 10.1109/ICAI52203.2021.9445249.

[14] A. Atrey, N. Narayan, S. Vijh, and S. Kumar, “Analysis of Breast Cancer using Machine Learning Methods,” in 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Jan. 2022, pp. 258–261. doi: 10.1109/Confluence52989. 2022.9734184.

[15] S. Jain and P. Kumar, “Prediction of Breast Cancer Using Machine Learning,” Recent Adv. Comput. Sci. Commun., vol. 13, no. 5, pp. 901–908, doi: 10.2174/2213275912666190617160834.

[16] P. P. Sengar, M. J. Gaikwad, and A. S. Nagdive, “Comparative Study of Machine Learning Algorithms for Breast Cancer Prediction,” in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Aug. 2020, pp. 796–801. doi: 10.1109/ICSSIT48917.2020.9214267.

[17] O. Tarawneh, M. Otair, M. Husni, H. Y. Abuaddous, M. Tarawneh, and M. A. Almomani, “Breast Cancer Classification using Decision Tree Algorithms,” Int. J. Adv. Comput. Sci. Appl. IJACSA, vol. 13, no. 4, Art. no. 4, 30 2022, doi: 10.14569/IJACSA.2022.0130478.

[18] E. Sugiharti, R. Arifudin, D. T. Wiyanti, and A. B. Susilo, “Integration of convolutional neural network and extreme gradient boosting for breast cancer detection,” Bull. Electr. Eng. Inform., vol. 11, no. 2, Art. no. 2, Apr. 2022, doi: 10.11591/eei.v11i2.3562.

[19] T. A. Assegie, R. L. Tulasi, V. Elanangai, and N. K. Kumar, “Exploring the performance of feature selection method using breast cancer dataset,” Indones. J. Electr. Eng. Comput. Sci., vol. 25, no. 1, Art. no. 1, Jan. 2022, doi: 10.11591/ijeecs.v25.i1.pp232-237.

[20] Z. Tasnim et al., “Classification of Breast Cancer Cell Images using Multiple Convolution Neural Network Architectures,” Int. J. Adv. Comput. Sci. Appl. IJACSA, vol. 12, no. 9, Art. no. 9, Sep. 2021, doi: 10.14569/IJACSA.2021.0120934.

[21] C. Kaul and N. Sharma, “High Accuracy Predictive Model on Breast Cancer Using Ensemble Approach of Supervised Machine Learning Algorithms,” in 2021 International Conference on Computational Performance Evaluation (ComPE), Dec. 2021, pp. 071–076. doi: 10.1109/ComPE53109.2021.9752254.

[22] Md. M. Islam, Md. R. Haque, H. Iqbal, Md. M. Hasan, M. Hasan, and M. N. Kabir, “Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques,” SN Comput. Sci., vol. 1, no. 5, p. 290, Sep. 2020, doi: 10.1007/s42979-020-00305-w.

[23] H. EL Massari, S. Mhammedi, Z. Sabouri, and N. Gherabi, “Ontology-Based Machine Learning to Predict Diabetes Patients,” in Advances in Information, Communication and Cybersecurity, Cham, 2022, pp. 437–445. doi: 10.1007/978-3-030-91738-8_40.

[24] H. El Massari, Z. Sabouri, S. Mhammedi, and N. Gherabi, “Diabetes Prediction Using Machine Learning Algorithms and Ontology,” J. ICT Stand., pp. 319–338, May 2022, doi: 10.13052/jicts2245-800X.10212.

[25] S. Srivastava, “Weka: A Tool for Data preprocessing, Classification, Ensemble, Clustering and Association Rule Mining,” Int. J. Comput. Appl., vol. 88, no. 10, pp. 26–29, Feb. 2014.

[26] H. El Massari, N. Gherabi, S. Mhammedi, Z. Sabouri, and H. Ghandi, “ONTOLOGY-BASED DECISION TREE MODEL FOR PREDICTION OF CARDIOVASCULAR DISEASE,” Indian J. Comput. Sci. Eng., vol. 13, no. 3, pp. 851–859, Jun. 2022, doi: 10.21817/indjcse/2022/v13i3/221303143.

[27] A. Khamparia and B. Pandey, “Comprehensive analysis of semantic web reasoners and tools: a survey,” Educ. Inf. Technol., vol. 22, no. 6, pp. 3121–3145, Nov. 2017, doi: 10.1007/s10639-017-9574-5.

An Intelligent Transport System in VANET using Proxima Analysis

Satyanarayana Raju K[1], Dr. Selvakumar K[2]
Department of Information Technology, Annamalai University, Chidhambaram, India

***Abstract*—There is no proper structure for Vehicular ad hoc networks (VANETs). VANET generates several mobility vehicles that move in different directions by connecting the vehicles and transferring the data between the source and destination which is very useful information. In this system, a small network is created with vehicles and other devices that behave like nodes in the network. Sometimes for better communication, VANET uses suitable hardware for improving the performance of the network. Reliability is one of the significant tasks that perform the needful operations and methods based on the conditions at a specific time. To disturb the VANETS, the attacker tries to hit the server and that causes damage to the server. This paper mainly focused on detecting the falsification nodes by analyzing the behavior of the models. In this paper, an improved intelligent transportation system (ITS) Proxima analysis is introduced to find the accurate falsification nodes. The proposed approach is the integration of KNN and RF with Proxima analysis. The main aim of the Proxima is to analyze the falsification nodes within the network and improve the mobility of the vehicles by sending source to destination without any miscommunication.**

Keywords—Vehicular Ad hoc Network (VANET); intelligent transportation system (ITS); KNN; RF

I. Introduction

VANETs is a very fast-growing field in wireless technology. VANET is considered as a sub-class of MANET in which the moving vehicles are considered as nodes or routers which are used to exchange the messages between the vehicles or access points. All the vehicles in this network are connected within the range of 100 to 900 meters by using 802.11p. This network will support both vehicles to a vehicle (V2V) and vehicle to infrastructure (V2I) for better communication within the network. The proposed approach ITS is used to increase road safety and provides a better travel experience for driver and passengers [1], [2].

Generally, VANET can increase traffic safety and effectiveness. Before starting the VANET network, privacy and security are two issues that are addressed before starting the network and data transmission [3]. In security, authentication is one of the significant tasks to provide privacy for the nodes present in the VANET network. With anonymous authentication, the network has to face challenges in verifying the vehicles in the network. This leads to a loss of data and communication between the vehicles [4]. This also fails to verify the huge data per second in VANETs. In VANETs, several advantages for routing protocols are identified based on the nature of the vehicle movements. A lot of research has been done on the issues of routing protocols such as scalability and reliability around the urban VANETs [5]. Nowadays Intelligent transportation systems (ITSs) are becoming more popular for connecting vehicles very efficiently and with better coordination. ITS's is the internal part of the VANET. This technique is mainly used to transfer the data between the nodes to improve security, efficiency, and reliability. VANETs are the sub-domain in the mobile ad-hoc networks (MANETs) and these are the integral parts of the ITSs. Strongly interconnected vehicles are used to sense the data and transfer the data based on location, traffic, environment, and urgent services [6].

VANET is showing attention because of several significant applications that are related to road safety and control of traffic. VANET mainly focused on improving road safety and controlling traffic. In smart cities, a lot of problems occur due to the heavy traffic due to the falsification nodes or vehicles. Falsification nodes create a lot of network damage that will lose the data which is transferred by the vehicles. By detecting the falsification nodes the Intelligent Transportation System (ITS) will improve the routing in the network. Based on the behavior of the vehicles falsifications vehicles are detected.

Fig. 1 shows the sample vehicles in the VANET's by using the SUMO simulator. All the yellow vehicles are normal vehicles that are having mobility. Fig. 2 shows the process of visualizing the VANET network by applying the proposed approach and its functionalities.

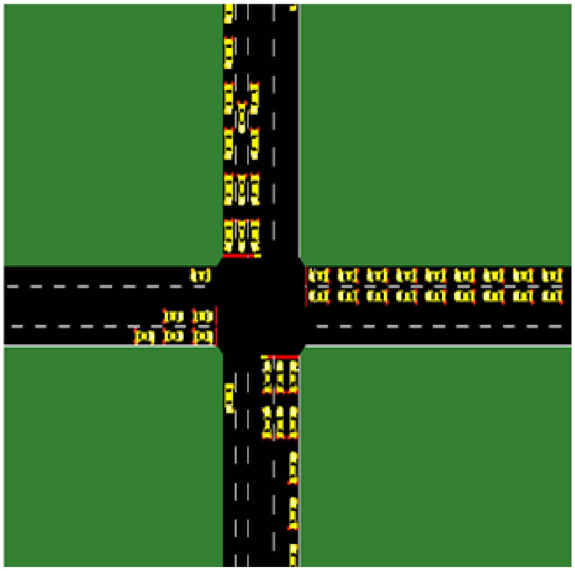


Fig. 1. Shows the Mobility of the Vehicles according to the Signals.

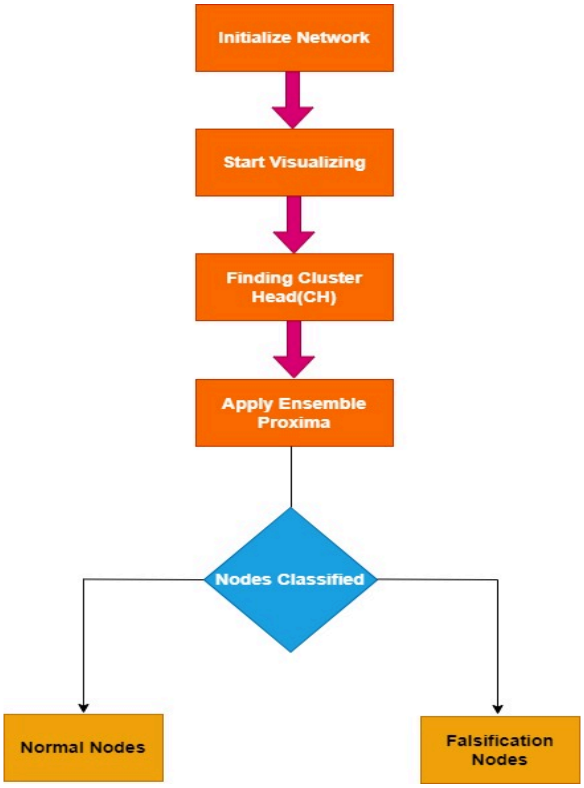


Fig. 2. Steps for Processing of Improved Intelligent Transportation System (ITS) Proxima Analysis.

II. Literature Survey

Over the last few years, Intelligent Transportation Systems (ITS) is showing an improvement in the movements of vehicles on the roads. The aim of the ITS is to provide better and more comfortable driving for the VANET vehicles that are present in the network by updating the information about the roads. Over the past many years, many researchers developed several approaches that are discussed in this section.

A. Ullah et al., [7] presented the location-based routing (LBR) protocol which is used to present the taxonomy. This approach is focused on analyzing the parked vehicles that are nearer to the junction for the selection of the path. This approach supports only less packet delivery, delay, and data transmission time and is more expensive. Abu Talib et al., [8] presented various security issues and solutions for the various issues and challenges in VANETs. This article also discussed the various types of attacks and various solutions that are implemented by solving several threats and shows the performance.

Abdul Quyoom et al., [9] proposed the ITS which is integrated with multifunctional system that collects a huge amount of data from several resources: Vision-Driven ITS (the sample data is collected from various visual sensors and utilized the vehicle and pedestrian detection); Multisource-Driven ITS (inductive, laser and GPS detectors); Learning-Driven ITS (this will reduce the collisions between the vehicles); and Visual-Driven ITS (this is used to find the abnormal traffic patterns and take required measures).

J. Cui et al., [10] introduced the adaptive approach that controls the traffic based on the communication among the cars. This system reduces the waiting time of the vehicles that are interacted with the decrease in queue length. To increase this system, clustering is adopted by utilizing the intersection of vehicles. By using this approach, the density of vehicles that are present in the cluster is calculated by using the clustering approach. The DBCV approach is used to increase the accuracy of the results. This approach is a combination of cluster and strategic diffusion methods which is utilized and collects the density information. Based on the movement and directions the clusters are formed within the region. By using the GPS and maps the direction of the vehicles is measured.

Saif Al-Sultan et al. [11] control the systems that are based on the other vehicles' data. Every vehicle is designed with a short communication device that controls the nodes that are combined with traffic lights. The node that controls this system plays the adaptive signal system. VANETs characteristics makes security and trust management as challenging issues. Different types of security threats and attacks persist in VANET [12].

III. Several Types of Attacks that Occur at the Simulation Time

Denial of Service (DoS) is an attack that can occur in the network [13]. There are two types of attackers, inside attacker and outside attacker. An inside attacker can jam the network by transmitting fake messages and stopping the network connections. The outside attacker continuously circulates the fake messages with fake signatures that use the bandwidth or other resources of a targeted vehicle. With this attack, VANET loses the ability to give services to the actual vehicles [14]. The main aim of this attack is to send the fake message to TSU and also to the actual vehicle to create a jam in the network.

The malicious nodes in the black hole attack [15], try to have the best route for the destination node and show that the data should transfer from this route by transferring the fake route information. A malicious node in this attack mainly drops or misuses the stopped packets without sending them to anyone [16]. This attack mainly creates the black hole area that creates the number of malicious vehicles and they are not interested to receive the messages from the actual vehicles [17].

Malicious vehicles in Wormhole attack [18], received the data at one point and replay it with another malicious vehicle by utilizing a wormhole high-speed link (tunnel) and data transfer from source to destination continuous by using malicious vehicles. This attack shows the huge impact on preventing finding the valid routes & menaces the security by transferring data packets. In this attack, the tunnel is used to broadcast the secret information by using two malicious vehicles.

The malicious vehicles in the sinkhole attack telecast the dummy routing information within the network [19]. This can easily attract the network traffic towards routing. This attack shows the huge impact on the network that complicates and reduces the performance by changing the data packets or dropping the packets [20]. The malicious vehicles in this network drop the data packets that are received from the authorized vehicle & telecast the fake routing info to the authorized vehicles on backside [21].

The malicious vehicles in the Sybil attack create a huge number of fake signatures that take overall control of VANET and insert the fake data in the network to threaten the legal vehicles. This attack will create the illusion among the multiple vehicles that creates a huge impact on the VANET network. This attack shows the huge impact on the network because of spoofing the signatures or places of other vehicles in the vehicular network [22]. This attack aims to create the fake identities of multiple vehicles and generates huge vehicles on the network [23].

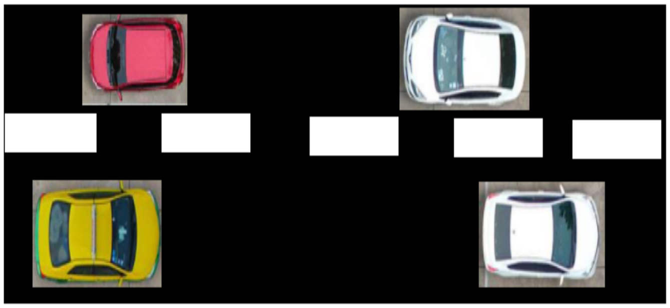


Fig. 3. In this Network Red Color Vehicle is considered as Malicious Node and other Vehicles are Normal Vehicles.

A. Random Forest (RF) for Detecting the Falsification Nodes

Random Forest (RF) is one of the hybrid approaches that follows the rule of bagging. It is an ensemble approach. Fig. 3 shows the normal nodes and malicious nodes by representing the malicious node with red color and normal vehicles are represented with white and yellow. This approach is a continuation of the decision tree (DT). DT is used to develop the information gain technique represented in Equation 1 and Equation 2.

To calculate the entropy value at every vehicle the equation (1) is given below:

$$\mathrm{E(s)} = \sum \mathrm{p_i log_2 p_i} \tag{1}$$

After entropy is calculated, the information gain is measured at every attribute to get the better decision for the vehicles.

$$\mathrm{Gain(S, A)} = \mathrm{Entropy(S)} - \frac{\Sigma_{\mathrm{v \in values(A)}} |\mathrm{S_v}|}{|\mathrm{S}|\mathrm{Entropy}} (\mathrm{S_v}) \tag{2}$$

Based on the entropy value the vehicles are classified according to the behavior.

The output of the random forest is given as an input to the KNN approach for filtering the falsification nodes according to the distance measure.

B. Applying KNN for RF Output Vehicles

KNN is one of the significant approaches used to classify the complex dataset. This is one of the better approaches for detecting the falsification nodes. Better training is given to the system for effective output. This approach analyzes the behavior of the vehicles in the real-time network. Based on the distance and direction the training generates better properties from the nodes. From the moving vehicles, the test runs are measured when the vehicles are in the same direction, different directions, and the location of vehicles are collected. For training, the dataset is utilized for the KNN model. If any abnormal behavior is identified among the vehicles, the distance and directions are identified as an input to the KNN classifier. The outputs of the KNN are represented as 0 or 1. 0 indicates the dangerous node and 1 indicates the safe. The KNN Algorithm Pseudocode:

```
1. Training and testing data is loaded
2. K-value is selected.
3. Every point in test data:
    ➢ Euclidean distance is used for training data points
      such as (a, b) and (-a, -b).

$d = \sqrt[2]{(a^2 + b^2}$                                   (3)

    ➢ List of Euclidean distances are stored and sorted.
    ➢ First k points are chosen.
    ➢ Based on the majority the class is assigned.

4. End
```

Above algorithm represents, K=5 i.e. five nearest neighbors are compared for every instance that requires to be classified. Hence it is known that the nearest vehicles are considered as normal vehicles and other vehicles are falsification vehicles based on the distance.

C. Measuring the Falsification Vehicles in the VANETs

In this paper, the proposed approach mainly focused on measuring the optimized routing by using proxima analysis. This approach considered the output of the KNN for proxima analysis. Proxima analysis is mainly focused on measuring the distance among the vehicles in the network. In this scenario, a proximity sensor plays the major role in detecting the distance among the vehicles and analyses the movement of vehicles. Here, the clustering head plays the major role to detect the abnormal vehicle.

- If the vehicle is having abnormal behavior then the vehicle drops or fakes the data packets received to create congestion in the network and mislead the vehicles and damage the malicious messages for their purpose.

- Truthful nodes forward the correct messages to the several nodes in the network and create the accurate messages for transmission.

- The main aim of this system is to monitor the behavior of the network. The monitoring process of the vehicles is called as “verifier” vehicles. Verifier is smaller or equal to Td compared with the Td of vehicle V, and this is placed inside the region z (V, Cluster Head (CH)). The intersection region is created for both vehicles named as V and CH.

By using these steps, the verifiers are monitored with V are send reports to CH [24]. The CH is equal to its transmission range and the V is obtained from the region and it is represented with Equation (4).

$$\text{Area (V)} = \text{TR(V)} - T_f(S_{max} - S_{min}) \tag{4}$$

Where,

S_{max} - Legal maximum speed of the vehicle. S_{min} - Legal minimum speed of the vehicle.

T_f – Packet Latency.

CH – Cluster Head.

Thus this will improve the routing of the vehicles by using proxima analysis.

Algorithm Steps for Proxima Analysis

```
Input: Nodes in the network Vehicles)
Output: Malicious Nodes (Vehicles)
Initialize variables N-Network, n-nodes or vehicles, r-region.
Step 1: $N_{mob}$-telecast the messages from all the vehicles.
Step 2: if ($N_{mob} \geq 2$)
Step 3: the mobility started it indicates presence of multiple nodes //Explains the status of the mobility
Else
Mobility not started.
Step 4: Calculate Euclidean distance by using Equation (3)
Step 5: arraylist [nearest nodes $N_n$] //Store the nearest nodes in arraylist.
Step 6: By using the given function nd_malv(Packet *p) the malicious vehicle is detected based on behavior.
if(mal==true)
{
drop(p,DROP_RTR_ROUTE_LOOP);
}
By showing the packet dropping rate the malicious nodes are identified.
Step 7: Malicious nodes are detected.
```

Table I shows the malicious and normal nodes. It is observed that in the existing approach there are missing nodes that are not considered by the existing approach. Compare with existing system approach the proposed approach detect very less missed vehicles.

TABLE I. SHOWING THE MALICIOUS AND NORMAL NODES

Total No of Nodes	Malicious Nodes		Normal Nodes	
	ES	PS	ES	PS
50	13	23	28	35
100	26	36	45	59
150	38	99	48	99
200	37	138	53	144

IV. Performance Metrics

A. Area under ROC Curve (AUC)

The overall proportion of true positives (malicious nodes accurately classified as malicious) compare with the proportion of false positives (not malicious and it is wrongly classified as malicious).

This is one of the efficient approaches that measures the following equation, Where t = (1 – specificity) and ROC (t) is sensitivity.

$$\text{AUC} = \int_0^1 \text{ROC(t)dt} \tag{5}$$

B. Accuracy

In machine learning, this is one of the significant metric that shows the overall accuracy of the data transmission and the performance of proposed approach. The overall accuracy is calculated by using below equation:

$$\text{Accuracy} = \frac{\text{TN+TP}}{\text{TP+TN+FP+FN}} \tag{6}$$

C. Precision

Precision is one of the significant factors in analyzing the results. High precision represents the low false positive rate.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{7}$$

D. Recall or Sensitivity

Recall is also one of the parameter to analyze the results. High recall relates to a low false negative rate.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{8}$$

E. Specificity

This measures the overall proportion of original negatives (Falsification Nodes), where the results are predicted as negative.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{9}$$

F. F1-Score

This is one of the measures that scores the weighted average of Precision and Recall. This will consider the false positives and negatives from the account. This will also show the uneven distributed classes.

$$\text{F1} - \text{Score} = \frac{2*(\text{Precision}*\text{Recall})}{(\text{Precision}+\text{Recall})} \tag{10}$$

Results and graphs were obtained with the following parametric values shown in Table II.

TABLE II. PARAMETRIC ENVIRONMENT

Metric	Values
Total No of Nodes	50-200 Nodes
Average Node Speed	50 m/s
Simulation Time	90-100 Sec

Table III and Fig. 4 show the performance of existing approaches by showing the AUC curve performance for 50, 100, 150 and 200 nodes. Among this the proposed approach Ensemble Proxima achieved the better results.

TABLE III. PERFORMANCE OF MACHINE LEARNING (ML) ALGORITHMS SHOWING AUROC

Total No of Nodes	KNN	KNN_RF	Ensemble Proxima
50	0.73	0.78	0.89
100	0.74	0.76	0.91
150	0.75	0.76	0.91
200	0.75	0.76	0.92

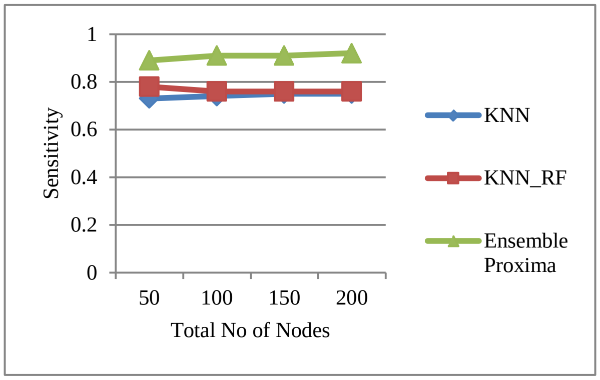


Fig. 4. Showing ROC for all the ML Algorithms.

Table IV, Fig. 5 and Fig. 6 show the performance of existing approaches by showing the Accuracy performance for 50, 100, 150 and 200 nodes. Among this the proposed approach Ensemble Proxima achieved the better results.

TABLE IV. PERFORMANCE OF MACHINE LEARNING (ML) ALGORITHMS SHOWING ACCURACY

Total No of Nodes	KNN	KNN_RF	Ensemble Proxima
50	84.49	89.85	93.03
100	85.12	88.89	94.12
150	86.45	89.89	95.12
200	85.56	88.98	95.78

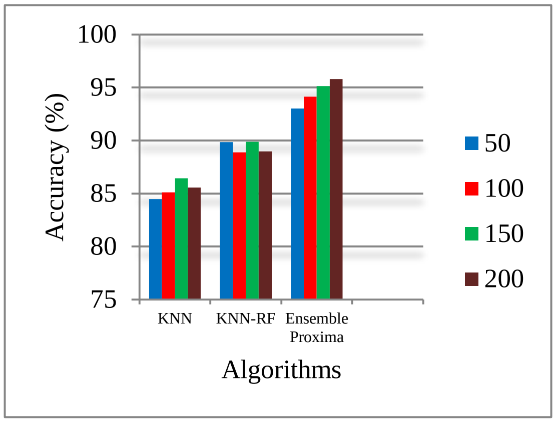


Fig. 5. Performance of Machine Learning (ML) Algorithms showing Accuracy.

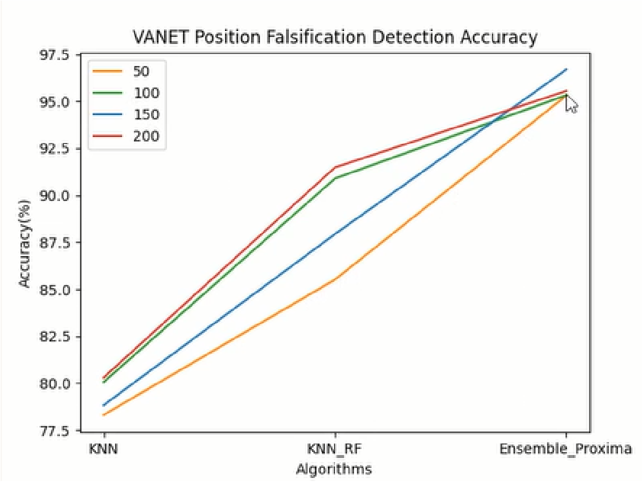


Fig. 6. Shows the Line Graphs based on Performance of Algorithms.

Table V, Fig. 7 and Fig. 8 explained about the performance of Existing and Proposed Algorithms. The performance is showing for precision parameter.

TABLE V. PERFORMANCE OF MACHINE LEARNING (ML) ALGORITHMS SHOWING PRECISION

Total No of Nodes	KNN	KNN_RF	Ensemble Proxima
50	83.79	88.15	94.63
100	84.32	87.89	95.42
150	85.15	88.89	96.32
200	84.16	89.98	96.78

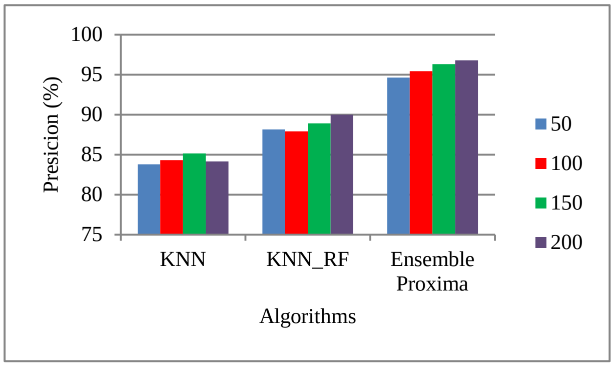


Fig. 7. Shows the Comparative Analysis of Existing and Proposed Algorithms for Precision.

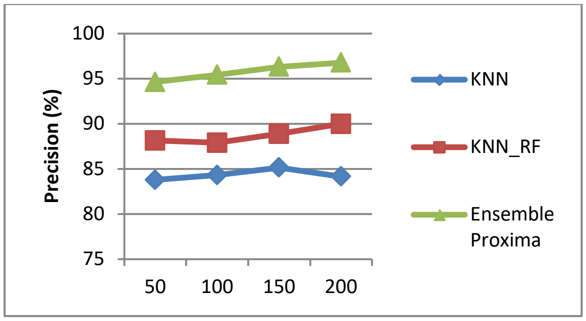


Fig. 8. Shows the Comparative Analysis of Existing and Proposed Algorithms for Precision using the Line Graphs.

Table VI, Fig. 9 and Fig. 10 show the performance comparison among the existing and proposed ML algorithms by showing the Recall. The performance is represented by showing bar graphs and line graphs.

TABLE VI. PERFORMANCES OF MACHINE LEARNING (ML) ALGORITHMS SHOWING RECALL

Total No of Nodes	KNN	KNN_RF	Ensemble Proxima
50	83.19	88.45	94.99
100	85.12	89.89	95.77
150	86.45	89.89	96.12
200	84.56	87.98	96.78

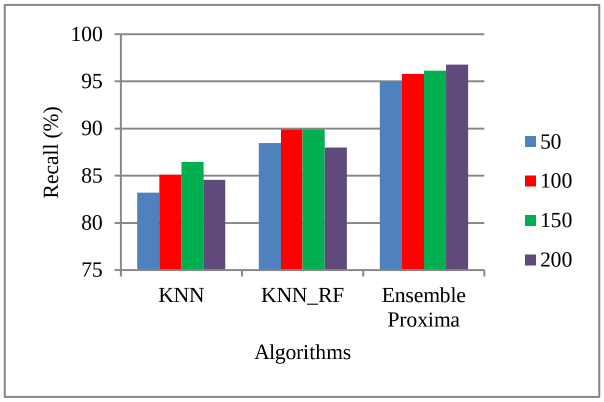


Fig. 9. Shows the Comparative Analysis of Existing and Proposed Algorithms for Recall.

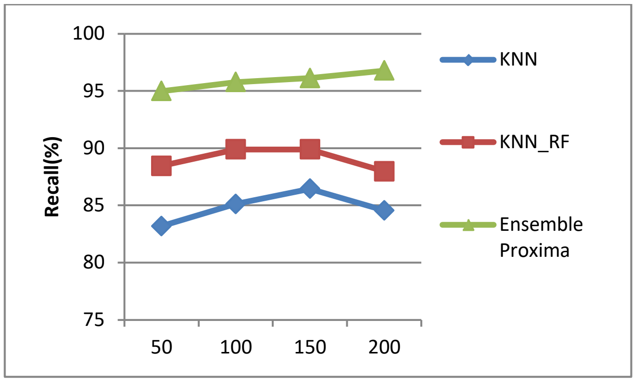


Fig. 10. Shows the Comparative Analysis of Existing and Proposed Algorithms for Recall using the Line Graph.

Table VII, Fig. 11 and Fig. 12 show the performance comparison among the existing and proposed ML algorithms by showing the F1-Score. The performance is represented by showing bar graphs and line graphs.

TABLE VII. PERFORMANCES OF MACHINE LEARNING (ML) ALGORITHMS SHOWING F1-SCORE

Total No of Nodes	KNN	KNN_RF	Ensemble Proxima
50	83.23	89.85	93.43
100	84.55	88.81	94.53
150	85.67	88.37	96.42
200	86.78	87.34	96.82

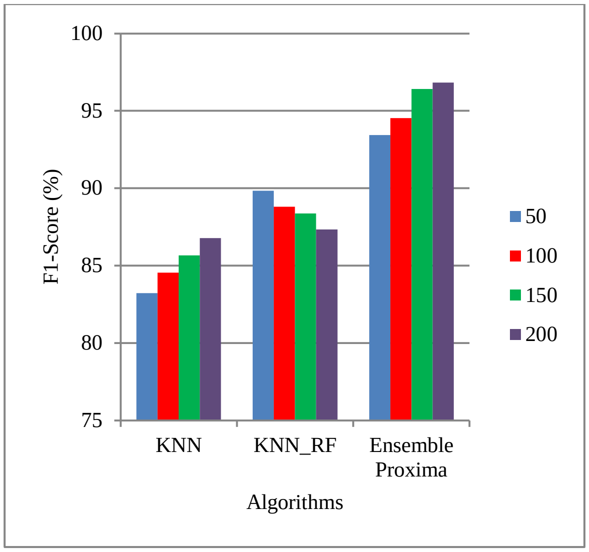


Fig. 11. Shows the Comparative Analysis of Existing and Proposed Algorithms for F1-Score.

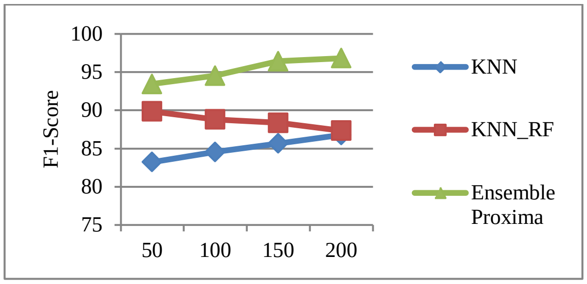


Fig. 12. Shows the Comparative Analysis of Existing and Proposed Algorithms for F1-Score using the Line Graph.

V. CONCLUSION

In this paper, the proposed approach focused on finding the falsification vehicles by using integrated Proxima analysis. The proposed approach is the combination of RF and KNN that helps to detect the falsification nodes based on the behavior of the vehicles. Falsification vehicles attack the network and cause breaking traffic rules such as over-speeding and wrong-way driving. It also provides a GUI which can be used by the traffic department to monitor roads and send help in case of an accident. The proposed method was cautiously evaluated in a traffic simulation environment, SUMO. The performance of the proposed approach has improved the performance in terms of the accuracy of AUROC. In future, an improved learning approach is combined with the various heuristic approaches to get the better detection of malicious nodes.

REFERENCES

[1] Javed Muhammad Noman et al., "VANET's Security Concerns and Solutions: A Systematic Literature Review," in Proceedings of the 3-rd International Conference on Future Networks and Distributed Systems (ICFNDS) ACM, pp. 1- 12, July 1-2, 2019.

[2] Abdul Quyoom, "Security Issues of Vehicular Ad Hoc Networks in OSI layers," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, ISSN: 2456-3307, vol. 2, no. 4, 2017.

[3] D. He, S. Zeadally, B. Xu and X. Huang, "An Efficient Identity-Based Conditional Privacy-Preserving Authentication Scheme for Vehicular

Ad Hoc Networks," in IEEE Transactions on Information Forensics and Security, vol. 10, no. 12, pp. 2681-2691, Dec. 2015, doi: 10.1109/TIFS.2015.2473820.

[4] M. Azees, P. Vijayakumar and L. J. Deboarh, "EAAP: Efficient Anonymous Authentication With Conditional Privacy-Preserving Scheme for Vehicular Ad Hoc Networks," in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 9, pp. 2467-2476, Sept. 2017, doi: 10.1109/TITS.2016.2634623.

[5] C. Cooper, D. Franklin, M. Ros, F. Safaei and M. Abolhasan, "A Comparative Survey of VANET Clustering Techniques," in IEEE Communications Surveys & Tutorials, vol. 19, no. 1, pp. 657-681, Firstquarter 2017, doi: 10.1109/COMST.2016.2611524.

[6] T. Chatterjee, R. Karmakar, G. Kaddoum, S. Chattopadhyay and S. Chakraborty, "A Survey of VANET/V2X Routing From the Perspective of Non-Learning- and Learning-Based Approaches," in IEEE Access, vol. 10, pp. 23022-23050, 2022, doi: 10.1109/ACCESS.2022.3152767.

[7] A. Ullah, X. Yao, S. Shaheen and H. Ning, "Advances in Position Based Routing Towards ITS Enabled FoG-Oriented VANET–A Survey," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 2, pp. 828-840, Feb. 2020, doi: 10.1109/TITS.2019.2893067.

[8] Abu Talib, Manar, et al., "Systematic literature review on Internet-of-Vehicles communication security," International Journal of Distributed Sensor Networks, ISSN: 1550147718815054, vol. 14, no. 12, 2018.

[9] Abdul Quyoom, MohdSaleem, MudasserNazar, Yusera Farooq Khan, "VANETs Applications, Challenges and Possible Attacks: A Survey, "International Journal of Advanced Research in Computer and Communication Engineering, ISO 3297:2007 Certified Vol. 6, Issue 7, July 2017.

[10] J. Cui, L. Wei, J. Zhang, Y. Xu, and H. Zhong, "An efficient message-authentication scheme based on edge computing for vehicular ad hoc networks," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 5, pp. 1621-1632, 2019. https://doi.org/10.1109/TITS.2018.2827460.

[11] Saif Al-Sultan, Moath M. Al-Doori, Ali H. Al-Bayatti, and HussienZedan, "A comprehensive survey on vehicular ad hoc network," Journal of Network and Computer Applications, vol.37, no. 1, pp. 380-392, 75 80 85 90 95 100 F1-Score (%) Algorithms 50 100 150 200 2014. https://doi.org/10.1016/j.jnca.2013.02.036.

[12] Z. Lu, G. Qu, and Z. Liu, "A survey on recent advances in vehicular network security, trust, and privacy, "IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 2, pp. 760-776, 2019. https://doi.org/10.1109/TITS.2018.2818888.

[13] Abdul Quyoom, Raja Ali and Devki Nandan Gouttam, "A Novel Mechanism of Detection of Denial of Service Attack (DoS) in VANET using Malicious and Irrelevant Packet Detection Algorithm (MIPDA)," in Proceedings of the IEEE International Conference on Computing, Communication and Automation (ICCCA2015), pp. 414- 419, 2015.

[14] Jafer, Muhammad, et al., "Secure Communication in VANET Broadcasting," ICST Transaction on Security Safety, vol.5, no.17, 2019.

[15] Karimireddy, T. and Bakshi, A., "A Hybrid Security Framework for the Vehicular Communications in VANET," in Proceedings of the International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 1929-1934, 2016.

[16] Satyanarayana Raju K, Dr. Selvakumar K, “Dynamic and Optimized Routing Approach (DORA) in Vehicular Ad hoc Networks (VANETs)”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 13, ISSUE No. 3, 2022. https://thesai.org/Publications/ViewPaper?Volume=13&Issue=3&Code=IJACSA&SerialNo=20

[17] Md Whaiduz-zaman, Mehdi Sookhak, Abdullah Gani, and Rajkumar Buyya, "A survey on vehicular cloud computing," Journal of Network and Computer Applications, vol. 40, pp. 325-344, 2014. https://doi.org/10.1016/j.jnca.2013.08.004.

[18] Sumra, Irshad Ahmed, Iftikhar Ahmad, HalabiHasbullah, and J-L. bin Ab Manan, "Classes of Attacks in VANET," in Proceedings of Saudi International Electronics, Communications and Photonics Conference (SIECPC), pp. 1-5, 2011.

[19] A. Festag. "Cooperative intelligent transport systems standards in Europe," Communications Magazine, IEEE, vol. 52, no.12, pp. 166-172, Dec. 2014. https://doi.org/10.1109/MCOM.2014.69799 70.

[20] Hussain Rasheed, Fatima Hussain, and Sherali Zeadally, "Integration of VANET and 5G Security: A review of design and implementation issues," Journal of Future Generation Computer Systems, pp. 843-864, 2019.

[21] Kumar Mr Kamal, and Rahul Malhotra, "Analysis of Sybil Attack Isolation Technique in VANET," International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 8, no. 5, pp.187- 192, May 2019.

[22] Singh Avinash et al., "Implementing Security Services in VANET Using Cryptography Based on Artificial Neural Network," Journal of Computer and Mathematical Sciences, vol. 10, no. 9, pp. 1573-1584, 2019.

[23] Karagiannis D. and Argyriou A., "Jamming attack detection in a pair of RF communicating vehicles using unsupervised machine learning," Vehicular Communications, vol.13, pp. 56-63, 2018. https://doi.org/10.1016/j.vehcom.2018.05.00 1.

[24] Safi, Q.G.K., Luo, S., Wei, C., Pan, L. and Yan, G., "Cloud-based security and privacy-aware information dissemination over ubiquitous VANETs," Computer Standards and Interfaces, vol.56, pp. 107- 115, 2018. https://doi.org/10.1016/j.csi.2017.09.009.

Virtual Reality, a Method to Achieve Social Acceptance of the Communities Close to Mining Projects: A Scoping Review

Patricia López-Casaperalta[1], Jeanette Fabiola Díaz-Quintanilla[2], José Julián Rodríguez-Delgado[3]
Alejandro Marcelo Acosta-Quelopana[4], Aleixandre Brian DuchePérez[5]
Universidad Católica de Santa María, Arequipa, Perú[1, 3, 4]
Universidad Católica San Pablo, Arequipa, Perú[2]
Universidad Privada Norbert Wiener, Lima, Perú[5]

***Abstract*—Background: Virtual reality (VR) technology is an effective, interactive and immersive type of communication since it produces greater interest and attention in the user, thereby allowing greater understanding and comprehension than with more traditional methods. On the other hand, not much information is known about the application of this novel technology in the context of social acceptance as far as the mining sector is concerned; our approach and methodology were based on scoping review (Prisma-SrC, Daudt et al., Arksey, and O'Malley). The research terms were also planned before, with the aim of carrying out three posterior screening levels, among which was the use of EndNote 20 and the PICO framework. Exhaustive research was carried out in nine databases. We obtained n=2 research articles of n=923 initially found, all of which went through three levels of filtering. The chosen articles were evaluated according to Hawker et al. 's methodological rigor, to be included in the review. This scoping review could be the starting point for a series of further investigations that would fill the gap in the literature on this topic, emphasizing experimental articles to confirm the impact of virtual reality technologies on the communities within the sphere of influence of a mining project.**

***Keywords*—Mining projects; social acceptance; virtual reality; interaction; mining communities**

I. Introduction

Nowadays, virtual reality (VR) is present in a wide range of research fields, such as medicine, aviation, military activities and space engineering, architecture, entertainment, leisure, and mining, etc. [1-3]. According to a study of twenty years of research on virtual reality, its basic characteristics have evolved considerably, improving the interactivity produced between the user and the system human-computer interaction (HCI), in addition to perfecting the feeling of immersion that the user perceives while in the virtual world [4, 5]. Furthermore, virtual reality can be performed with exceptionally high quality, reproducing images and changing data three-dimensionally [6] providing a highly realistic simulation of the real world [7] and its continued development would facilitate the possibility of applying powerful new tools to influence the user's perceptual psychology [5].

On the other hand, social acceptance is related to the degree of tolerance that a community surrounding a mining project has for said project and whether or not the mining company can carry out its operations. This is considered very important since, without prior acceptance, the Social License to Operate (SLO) [8] would not be issued and, therefore, it would not start operations [9], generating economic losses to the company and the loss of benefits delivered to communities by mining companies [10] (mining taxes and royalties, direct and indirect employment).

Currently, energy and mining companies tend to justify their projects to the surrounding communities through the use of traditional methods and tools [11] which are usually represented by 2D images, plans and static drawings [6]. However, this does not guarantee either effective or convincing transmission of information with the community, which could create tension and conflicts due to misunderstandings among the stakeholders [10, 12]. A digital visualization of a landscape could stimulate public participation and allow for broader input regarding questions related with industrial projects [7].

The objective of this scoping review was to carry out an exhaustive investigation with the aim of determining the opportunities that the application of virtual technology could provide in support of mining projects and, based on the eventual findings, to identify gaps in the literature to help the planning and implementation of future research [13, 14].

A. Virtual Reality (VR)

This is defined as the hardware-software binomial, which works in the interaction with the user through effective synchronous communication [7], and it has the capacity for isolation from real life and immersion in a virtual life [15,16]. Virtual reality could also solve even more complex problems [15].

Immersive Virtual Reality (IVR) is defined directly with the concept of emphasizing the act of being present in the virtual environment, where the quality of immersion is directly proportional to the effectiveness of the experience [16], which may raise public awareness more effectively and result in a higher rate of acceptance of diverse projects [7]. In more simple terms, in order to obtain optimum results, you must work under the Triangle matrix of Virtual Reality (Interaction-Imagination-Immersion) [2, 3].

II. Materials and Methods

For the purposes of this study, a PRISMA extension methodology for scoping reviews was employed (PRISMA-SrC) [13], as well as the methodology of Arksey and O'Malley [17] and Daudt et al. [18]. Furthermore, to provide added value to the scoping review, step six by Arksey y O'Malley [17] who considered the exercise of additional inquiry was carried out (Daudt et al. [18] y Levac et al. [19] discuss and consider it mandatory). As argued by Daudt et al. [18], too many people in a research group involves a risk that there will be several diverse interpretations during a scoping review; so it was decided to carry out this article with only four participating researchers.

A. Search Plan

First of all, before search criteria were developed, for which a variety of compiled tasks were carried out in August 2021 from nine distinct online databases: Scopus, Web of Science, IEEE, EBSCOhost and ScienceDirect ACM Digital Library, ERIC, SciELO, Google Scholar. The researcher (KV) supervised the means of obtaining data, as well as the use of Boolean operators and thesauri. The following link shows Table I detailing the databases and their search criteria.

TABLE I. Search Criteria According to Database used

Electronic Databases	Search Strategy
IEEE	(("All Metadata":interact* OR "All Metadata":acceptance) AND ("All Metadata":publics OR "All Metadata":public OR "All Metadata":communit*) AND ("All Metadata":project* OR "All Metadata":compan* OR "All Metadata":mining) AND ("Document Title":"virtual reality" OR "Document Title":VR))
ScienceDirect	Title, abstract, keywords: (interact OR acceptance) AND (public OR publics OR community) AND (project OR company OR mining) Title: ("virtual reality" OR VR)
Scopus/Web of Science	Title, abstract, keywords: (interact* OR acceptance) AND (public OR publics OR communit*) AND (project* OR compan* OR mining) Title: ("virtual reality" OR VR)
EBSCOhost	Title, abstract, keywords: (interact* OR acceptance) AND (public OR publics OR communit*) AND (project* OR compan* OR mining) NOT (patient* AND rehabilitation) Title: ("virtual reality" OR VR)
ACM Digital Library	[[All: interact*] OR [All: acceptance]] AND [[All: public] OR [All: publics] OR [All: communit*]] AND [[All: project*] OR [All: compan*] OR [All: mining]] AND [[Publication Title: "virtual reality"] OR [Publication Title: vr]]
ERIC	(interact* OR acceptance) AND (public OR publics OR communit*) AND (project* OR compan* OR mining) Title: ("virtual reality" OR VR)
SciELO	((interact* OR acceptance) AND (public OR publics OR communit*) AND (project* OR compan* OR mining)) AND (title: (("virtual reality" OR VR))))
Google Scholar	CON TODAS LAS PALABRAS (interact acceptance public community project company mining) CON LA FRASE EXACTA (intitle:virtual intitle:reality)

The search criteria were limited to journals, conference papers, and reviews, with only publications in English from 2010 to 2021 taken into account. Subsequently, the search results were imported into EndNote 20 software, which is a reference manager that allowed us to proceed with the second filter. Later, the screening process was performed, and was based on the PICO framework (Population-Intervention or Exposure-Comparison-Results) [14].

B. Screening Criteria

The second review was carried out by two researchers (AA & JR), who used EndNote 20 to delete the duplicated articles. Additionally, the researchers identified, verified, and removed irrelevant articles manually. The third filter was performed under the PICO framework [14], with the aim of finding articles related to this research project. Each aspect of the eligibility criteria is detailed in Table II. Screening was carried out by performing manual searches in which all researchers participated (AA, KV, MB, PL, JR).

TABLE II. Type of Screening used for Selection of Studies

Electronic Databases	Search Strategy
Population Communities in general.	Homogeneous communities, Clients, Patients
Intervention or Exposure Virtual Reality Technology (VR), Immersive Virtual Reality Technology (IVR), Cinematographic Virtual Reality Technology (CVR)	Non-Immersive Reality Technology (NIVR), Augmented Reality (AR) Technology, Smartphone Applications, 2D simulation technology.
Comparison Pre-test, Post-test.	If it lacks any evidence mentioned in the Inclusion Criterion, then it is discarded.
Results If it lacks any evidence mentioned in the Inclusion Criterion, then it is discarded.	If the approach is not related to Social Acceptance, then the article is discarded.

C. Methodological Rigor

The evaluation proposed by Hawker et al. helps to enhance the quality of this article in order for it to be useful and subsequently disseminated for future research [18]. The research articles that passed the three levels of filtering were evaluated according to the methodology of Hawker et al. [20] by two reviewers (AA, JR), according to the following evaluation criteria: "Good", "Fair", "Poor", and "Very Poor". This design framework allows us to evaluate a wide range of research methods. For this reason, we considered the parameters of Appendix D regarding to the methodological rigor analysis [20].

III. Results

Table III, which contains the results of the evaluation of methodological rigor [20], resulting in the approval of the quality of both articles for the development of this scoping review.

TABLE III. RESULTS OF EVALUATION OF METHODOLOGICAL RIGOR APPLIED IN THE SELECTED STUDIES

Methodological Rigor Analysis	**Worth a thousand words: Presenting wind turbines in virtual reality reveals new opportunities for social acceptance and visualization research**	**The 3D Immersive Virtual Reality Technology Use for Spatial Planning and Public Acceptance**
1. Abstract and title: Did they provide a clear description of the study?	Good	Good
2. Introduction and aims: Was there a good background and clear statement of the aims of the research?	Good	Good
3. Method and data: Is the method appropriate and clearly explained?	Good	Good
4. Sampling: Was the sampling strategy appropriate to address the aims?	Good	Good
5. Data analysis: Was the description of the data analysis sufficiently rigorous?	Good	Good
6. Ethics and bias: Have ethical issues been addressed, and what has necessary ethical approval gained? Has the relationship between researchers and participant been adequately considered?	Good	Good
7. Results: Is there a clear statement of the findings?	Good	Good
8. Transferability or generalizability: Are the findings of this study transferable (generalizable) to a wider population?	Good	Good
9. Implications and usefulness: How important are these findings to policy and practice?	Good	Good
General Perception	Good	Good

Using the search criteria, n=923 articles were found, possibly related to the aim of the present research. Then, the information was screened until n=2 articles, according to their methodological rigor evaluation, as shown in Fig. 1.

A. Research Summary

An article selected and included in the present investigation emphasized its experimental objective by demonstrating that the use of technologies such as Immersive Virtual Reality (IVR) can be of great benefit to companies, mainly to provide a competitive advantage in the market [21]. The study was based on the range and variety of problems that wind energy projects go through before and during the construction process and it was demonstrated that the methods used to establish communication with the public are usually ineffective. Two tests were carried out before and one after the experimentation with 70 participants, where 50% were from the male population and the remainder from the female population. Additionally, the article highlights that 21.4% of the participants were over 28 years of age. The research findings showed an 8.6% increase in the post-test concerning the change in the participants' attitudes towards the proposed technology's general acceptance. Likewise, 40% of participants changed their perception in the pre-test, since the post-test results showed that 62.8% considered the project harmful to animals and the environment. On the other hand, 97.1% of the participants affirmed that immersive virtual reality technology was essential to change their attitude about energy projects, helping to better understand what they wanted to teach. On the other hand, only 38.6% of the participants considered 2D technology significant. The results found in this research show that the existing gap in communication between companies and people is reduced when implementing 3D IVR. It was also revealed that the use of this technology should play a more significant role in companies since it can unify the level of understanding between participants who are experts in a subject and those with less experience [21].

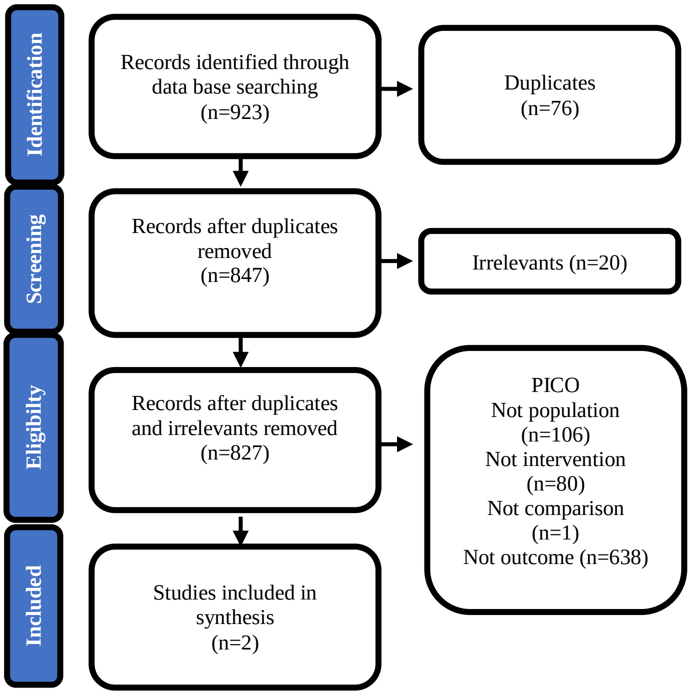

Identification
Screening
Eligibilty
Included
Records identified through data base searching (n=923)
Duplicates (n=76)
Records after duplicates removed (n=847)
Irrelevants (n=20)
Records after duplicates and irrelevants removed (n=827)
PICO
Not population (n=106)
Not intervention (n=80)
Not comparison (n=1)
Not outcome (n=638)
Studies included in synthesis (n=2)

Fig. 1. Filter Stages and Selection of Studies.

Another article developed in the United States [11], investigates how the interaction, between the study population and the wind turbine represented by cinematographic virtual reality (CVR), influences the expectations and perceptions of wind energy projects. Specifically, it studies the relationship between personal appreciation of energy projects and changes in opinions after interacting with CVR. The population was recruited for the student-based experiment, all of whom were eligible to participate in the project, and the study group only knew about the subject from the mentions made by the recruiter, so it was considered a "naive population for the purposes of the study" (For this reason the authors include this article as part of the final filter of the quality review). A total of 101 students (47 women and 54 men) carried out an entrance survey to determine the population’s knowledge of before energy technologies. The questions are divided into seven categories: support and opposition, personal feeling, overall impact, impact on wind farms, attitude, knowledge, and visual and acoustic questions. 74% of the participants had never seen a wind turbine before, 54% had already experimented with virtual reality at some point in their lives, and 30% of the participants felt they knew enough about wind energy before experimentation. Afterwards, the participants underwent a test phase where they spent two minutes immersed in a virtual reality video to acclimatize and prepare the population. Participants who had never seen a wind turbine, and who did not know about it, were the ones who presented the most significant changes in their attitudes and opinions after the experiment. The experience turned out to have a high level of immersion and ecological validity, reducing erroneous beliefs about the different impacts of wind turbines- It can be assumed, therefore, that virtual reality can be profitably used to communicate more clearly the development of wind projects to the public in general, especially if the population's has poor knowledge of the subject. It goes on to propose that developers and politicians take such technological advances into account these since these are new tools which enable more effective interaction(s) with the population. A higher quality of image and audio in the presentation of a proposed project through virtual reality will help the public become more aware of and better understand the real impacts of the project and therefore will have greater influence in its acceptance [11].

You can see Table IV specifying the main characteristics of both articles.

TABLE IV. OUTSTANDING CHARACTERISTICS OF SELECTED STUDIES

Article	**Worth a thousand words. Presenting wind turbines in virtual reality reveals new opportunities for social acceptance and visualization research**	**The 3D Immersive Virtual Reality Technology Use for Spatial Planning and Public Acceptance**
Research Aim/ Objective	Examine how a multimodal and cinematographic virtual reality (CVR) experience with a wind turbine affects the expectations and perceptions of wind power projects. Specifically, we investigated the relationship between personal opinions about (renewable and non-renewable) energy sources, and variations in perceived impacts.	The objective is to evaluate the use of 3D virtual reality to support innovative SMEs in addressing issues related to spatial planning and corresponding public acceptance.
Outcomes	Students completed the experiment and approximately half (54%) had previously experienced virtual reality. Three quarters of the participants (74%) had never seen a wind turbine up close. About a third of all participants (30%) felt they were well informed about wind energy prior to the study. The CVR experience provided a high level of immersion and ecological validity. The CVR experience corrected previous erroneous beliefs regarding the expected visual and acoustic impacts of wind turbines.	There was a significant change in the perception of the participants before and after the session. For example, there was an increase of approximately 8.6% of the participants who positively changed their perception about the appearance of wind turbines. It also increased by almost 40% of participants who considered wind turbines as "bird friendly". The preference of people in 97.1% with respect to immersive virtual reality was also observed, tripling the result obtained in 2D technology.
Conclusions	In many cases, participants who had not seen a wind turbine before or who had felt less informed had greater changes in attitudes and opinions after the CVR experience. The participants felt immersed in the scene. Most of the participants considered that the wind turbine had a neutral or positive impact on the scene, although we cannot reject the null hypothesis that men and women reported feeling equally knowledgeable about wind energy. After the CVR experience, visual concerns about offshore wind farms having negative impacts on the landscape decreased. Taken together, these results suggest that CVR can be leveraged to more clearly communicate wind project development plans to the general public, especially when prior audience knowledge and experience with wind turbines is limited. We recommend that developers and managers, as well as politicians consider the proposed multimodal virtual experiences with wind turbines as a new tool to interact with the public. Accurately representing the visual and auditory characteristics of a proposed project through immersive visual and auditory simulations can help the public understand what to expect regarding project impacts, potentially allowing communities and developers to interact with one another more significantly.	The gap produced between companies and the public can be drastically reduced due to the implementation of 3D IVR As an SMEs. McCamley's management first reported a positive impact on the implementation of innovative 3D IVR within both its internal and external decision-making process. Internally, the company enriched its knowledge of the importance of technology in supporting engineering and management decisions, promotional events, and discussions with investors. In addition, dynamic simulation in the 3D IVR environment provides the public and the authority with a better understanding of the product. This allowed public support for one of the key product features; that of it being bird friendly. The commercial importance of the brand and color becomes a clear management decision because it can broaden public acceptance in terms of turbine appearance and shorten the payback

Article	**Worth a thousand words. Presenting wind turbines in virtual reality reveals new opportunities for social acceptance and visualization research**	**The 3D Immersive Virtual Reality Technology Use for Spatial Planning and Public Acceptance**
		period for investors. The company endorses the value of technology by integrating immersive 3D virtual reality into diverse management processes.
Methods and Methodology	The entry survey asked participants a series of questions about their opinions on different energy technologies in general and wind energy in greater detail. These questions were divided into seven categories: support and opposition, personal feeling, general impact, wind farm impact, attitude, knowledge, and visual and acoustic questions. During the testing phase, participants spent two minutes in a virtual museum in order to (a) acclimatize them to virtual reality, and (b) prepare them to put themselves in their place. A group asked about the visual and acoustic elements of the scene, including the volume and how well the wind turbine fit into the scene. Another group evaluated participants' previous experience with virtual reality and self-reported level of immersion in the scene using selected elements from Witmer and Singer.	A prior survey was conducted to measure the perception of 70 participants on wind energy projects. 10 sessions were held, and each session involved the participation of 5 to 9 people. After that the participants experienced 3D IVR, once the participants finished the experience, they carried out subsequent questionnaires to determine their changing perceptions.
Study Population	Participants were recruited between November 2018 and May 2019 and were compensated ($ 10 or course credit) upon completing the experiment. Participants were naive of the purpose of the study, although recruitment materials did mention wind energy. A total of 101 (47F, 54M; Age 19.5 years, SD Age 1.44 years) students completed the experiment and approximately half (54%) had previous experience of virtual reality. All students were eligible to participate in the study.	70 participants, of whom 50% were men, and 50% women. 21.4% were over 28 years old, 64.3% were participants with a history.

B. Stakeholder Views

According to a professional in mining operations, virtual reality would be an improved alternative when dealing with affected communities. This technology can surpass conventional methods in supporting mining projects within the surrounding areas and communities. Consideration should also be given to providing information and training through the application of virtual reality technology since it allows real time simulations similar to real experience, reducing risks, negative impacts and operating costs. He also commented that human-computer interaction (HCI) offers adequate retention, similar to mining training. A second professional (GP) said that this technology is used in shovels and mining trucks, claiming that it was a success. He argues that everyone in large mining projects invests in technology and virtual reality. He proposed to perform visual modeling of wind direction so that members of the surrounding communities can understand how controls are carried out in operational units such as blasting and loading, which are generators of particulate material. HS, involved in the agricultural sector further assured that the use of virtual reality to support mining projects would show the proposed execution of said projects would be undertaken in a more refreshing and convincing way.

IV. Discussion

Individual differences were found in aspects of the selected articles. This does not mean that they have a low methodological rigor [16], but these differences can make them more complementary. For example, the type of virtual reality technology used by Cranmer et al. [11] was different from that of Abulrub et al. [21], cinematographic (CVR), and immersive (IVR) respectively. Despite such differences, both results were positive, because these technologies are effective for their purpose, so it can be logically assumed that human-computer interaction is effective [22]. The interaction between virtual reality technology and the user has the most significant impact on those who have little or no experience with said technology [11]. Therefore, it is believed that virtual reality could be efficient in the communities lying within the sphere of influence of a mining project since most of them do not have any (previous) access to Information and Communication Technologies (ICT) [23], as highlighted in a 2018 study which stated that the use of virtual reality technology could be considered to be the main and most effective means of communication with the public due to its immersive capacity [7].

Due to the few articles found in the databases specified in the methodology, it is possible to argue that this research could be the first scoping review that refers to the use of virtual reality technology to generate a new opportunity for the social acceptance of mining projects. When performing an in-depth and exhaustive investigation into the databases mentioned in the methodology, very few articles were found related to the use of virtual reality for the social acceptance of mining projects. Indeed, no articles at all were found which were directly associated with the last variable. As a consequence, it was decided to analyze the experience with wind-energy projects, since both these and mining projects have a common problem: dealing with social acceptance [10] and reducing the gap in understanding and communication between the project and local communities [24]. Although it is possible to extract valuable information on the social acceptance of renewable energy (RE) projects, the appreciation is not the same since their operations do not have the same impact [25], For this reason, more experimental research is needed to know more precisely what the appreciation of the communities towards mining projects is and how HCI with virtual reality could reduce the understanding-communication gap [2, 21]. It is worth mentioning that Cranmer et al. [11], carried out his experimentation with a student community. Nevertheless, it was decided to include this article as part of the scoping review, since it shares the technological objective of this research and the results of the opinions were similar to those of the population from Massachusetts.

According to the research included for review and analysis, conventional 2D methods are still used currently to support projects [6], although it is stated that these traditional methods do not have enough interaction, nor the same understanding produced by experimenting with virtual reality technology [11,21], even though these technological methods are extraordinarily superior to traditional ones [1-3,6,11,15,21-22].

ICTs have been continuously improving over the years, and these tools can concretely and effectively support the teaching and learning process [26] that occurs when presenting a mining project to a specific population to achieve its acceptance and implementation. The use of ICT can be used by the personnel in charge of informing the public about the mining project, since these are flexible, accessible, affordable and without limitations in terms of time and space, so that the population with the mine's sphere of influence can have a good understanding of the mine's operations, activities and intentions [27].

The professionals' opinion provide us with support for the benefit acquired by using virtual reality technology as well as providing us with additional ideas to support mining projects, such as the idea proposed by (GP) on the creation of a visual model of wind management. In that way, the impact minimization plan [28] will practically be shown to residents, seeking to demonstrate that the benefit that the community will obtain by accepting the project will be higher than the cost of the impacts generated, as long as the corresponding legislation of the country where it would operate [10]. The idea of a professional in mining operations about the use of virtual reality technology, would surpass the understanding and understanding of conventional methodologies, coinciding with the opinion of (HS) an agricultural representative of a community within the sphere of a mining project, who believes it would improve understanding and, likewise, reduce the uncertainty of the local population [11,21,29], (HS) affirming that it would be a different and novel way of explaining the aspects of a mining project regarding its benefits, stages, operational processes and mitigation plan specified in the Environmental Impact Study. In this case, the farmer's point of view is highly important, since a social conflict between mining and agriculture is often generated.

The authors believe that the use of virtual reality can improve the planning of a mining project with the participation of the surrounding community, thereby ensuring a common benefit and reaching mutual agreement, as specified by Yu et.al [7], where public participation makes wind projects more acceptable.

V. Conclusion

The accessibility to the UCSM virtual library allowed us to do our research freely without any restrictions regarding different international databases, such as those mentioned in the methodology. Also, the opinion that was obtained from the interested parties was considered a strength. A limitation was the scarcity of existing research regarding the interaction of the local community with virtual reality in the context of social acceptance of mining projects. The current situation (SARS-CoV-2) is also considered as a difficulty to obtain more data or opinions from professionals and people close to a mining project.

We conclude that we have little scientific research information regarding the use of virtual reality in the interaction of neighboring communities for the social acceptance of mining projects.

Based on the experimental experiences from the selected articles, we conclude that immersive virtual reality (IVR) as well as cinematographic virtual reality (CVR) provides better communication and understanding between the user and the company in comparison to the traditional methods, improving the acceptance percentage of the proposed project.

The authors trust that this scoping review article will be the starting point to begin the experimental research and the future systematic reviews to contribute the use of virtual reality technology for the development of effective and interactive communication.

Acknowledgment

The authors declare that there is no conflict of interest. We thank the Universidad Católica de Santa María for providing us the easy access to scientific databases. We thank the professionals consulted in the mining sector: Eng. Milton Trujillo Espinoza, Eng. Giancarlo Pinto Guerra, and the farmer Helbert Samalvides for their important points of view on this research.

References

[1] Z. Xiaoqiang, W. An, and L. Jianzhong, "Design and application of virtual reality system in fully mechanized mining face", Procedia Engineering, vol. 26, pp. 2165-2172, 2011.

[2] Z. Hui, "Head-mounted display-based intuitive virtual reality training system for the mining industry", International Journal of Mining Science and Technology, vol. 27, pp. 717-722, 2017.

[3] L. Jun, and L. Guo-bin, "Research on key techniques of virtual reality applied in mining industry", Journal of Coal Science & Engineering, vol. 15, pp. 445-448, 2007.

[4] S. Correia, J. Guerreiro, and F. Ali, "20 years of research on virtual reality and augmented reality in tourism context: A text-mining approach", Tourism Management, vol. 77, 104028, 2020.

[5] E. Crofton, C. Botinestean, M. Fenelon, and E. Gallagher, "Potential applications for virtual and augmented reality technologies in sensory science", Innovative Food Science and Emerging Technologies, vol. 56, 102178, 2019.

[6] J. Toraño, M. Menéndez, M. Gent, and I. Diego, "A finite element method (FEM) – Fuzzy logic (Soft Computing) – virtual reality model approach in a coalface longwall mining simulation", Automation in Construction, vol. 17, pp. 413-424, 2008.

[7] T. Yu, H. Behm, R. Bill, and J. Kang, J., "Validity of VR Technology on the Smartphone for the Study of Wind Park Soundscapes", Geo-Information, vol. 7, pp. 1-13, 2018.

[8] M. Boateng, and K. Awuah-Offei, "Agent-based modeling framework for modeling the effect of information diffusion on community acceptance of mining", Technological Forecasting & Social Change, vol. 117, pp. 1-11, 2017.

[9] S. Leena, U. Karina, and L. Jungsberg, "Social license to operate in the frame of social capital exploring local acceptance of mining in two rural municipalities in the European North", Resources Policy, vol. 64, 2019, 64, 101498.

[10] A. Zhang, and K. Moffat, K., "A balancing act: The role of benefits, impacts and confidence in governance in predicting acceptance of mining in Australia", Resources Policy, vol. 44, pp. 25-34, 2015.

[11] A. Cranmer, J. Ericson, A. Broughel, B. Bernard, E. Robicheaux, and M. Podolski, "Worth a thousand words: Presenting wind turbines in virtual reality reveals new opportunities for social acceptance and visualization research", Energy Research & Social Science, vol. 67, pp. 1-10, 2020.

[12] M. Rodrigues, and L. Mendes, "Mapping of the literature on social responsibility in the mining industry: A systematic literature review", Journal of Cleaner Production, vol. 181, pp. 88-101, 2018.

[13] Y. Zhang, H. Liu, H., S. Kang, and M. Al-Hussein, "Virtual reality applications for the built environment: Research trends and opportunities", Automation in Construction, vol. 118, 103311, 2020.

[14] A. Tricco, E. Lillie, W. Zarin, K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. Peters, T. Horsley, L. Weeks, S. Hempel, E. Akl, C. Chang, J. McGowan, L. Stewart, L. Hartling, A. Aldcroft, M. Wilson, C. Garritty, S. Lewin, C. Godfrey, M. Macdonald, E. Langlois, K. Soares-Weiser, J. Moriarty, T. Clifford, and S. Straus, "PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation", Annals of internal medicine, vol. 169, pp. 467-473, 2018.

[15] J. Bellanca, T. Orr, W. Helfrich, B. Macdonald, and J. Navoyski, "Developing a Virtual Reality Environment for Mining Research", Mining, Metallurgy & Exploration, vol. 36, pp. 597-606, 2019.

[16] J. Cummings, and J. Bailenson, "How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence", Media Psychology, vol. 19, pp. 37-41, 2015.

[17] H. Arksey, and L. O'Malley, "Scoping studies: towards a methodological framework", International Journal of Social Research Methodology, vol. 8, pp. 19-32, 2005.

[18] H. Daudt, C. Van Mossel, and S. Scott, "Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework", BMC Medical Research Methodology, vol. 13, pp. 1471-2288, 2013.

[19] D. Levac, H. Colquhoun, and K. O'Brien, "Scoping studies: advancing the methodology", Implementation Science, vol. 5, pp. 1-9, 2010.

[20] S. Hawker, S. Payne, C. Kerr, M. Hardey, and J. Powell, "Appraising the Evidence: Reviewing Disparate Data Systematically", Qualitative Health Research, vol. 12, pp. 1284-1299, 2015.

[21] A. Abulrub, K. Budabuss, P. Mayer, and M. Williams, "The 3D Immersive Virtual Reality Technology Use for Spatial Planning and Public Acceptance", Procedia - Social and Behavioral Sciences, vol. 75, pp. 328-337, 2013.

[22] Y. Lau, Y. Tang, I. Chan, A. Ng, and A. Leung, "The deployment of virtual reality (VR) to promote green burial", Asia Pacific Journal of Health Management, vol. 15, i403, 2020.

[23] G. Caspary, and D. O'Connor, Providing low-cost information technology access to rural communities in developing countries: What works? What pays? Washington DC: OECD Development Centre Working Papers, 2013.

[24] L. Mercer-Mapstone, W. Rifkin, W. Louis, and K. Moffat, "Company-community dialogue builds relationships, fairness, and trust leading to social acceptance of Australian mining developments", Journal of Cleaner Production, vol. 184, pp. 671-677, 2018.

[25] B. Walsh, S. Van der Plank, and P. Behrens, P., "The effect of community consultation on perceptions of a proposed mine: A case study from southeast Australia", Resources Policy, vol. 51, pp. 163-171, 2017.

[26] P. López, J. Rodríguez, A. Acosta, and M. Berrios, "Analysis from the student perspective on the implementation of learning technologies in mining engineering", CEUR Workshop Proceedings, vol. 2555, pp. 268-277, 2019.

[27] C. Shen, and J. Ho, "Technology-enhanced learning in higher education: A bibliometric analysis with latent semantic approach", Computers in Human Behavior, vol. 104, 106177, 2019.

[28] S. Que, K. Awuah-Offei, and V. Samaranayake, "Classifying critical factors that influence community acceptance of mining projects for discrete choice experiments in the United States", Journal of Cleaner Production, vol. 87, pp. 489-500, 2015.

[29] S. Rourke, "How does virtual reality simulation compare to simulated practice in the acquisition of clinical psychomotor skills for pre-registration student nurses? A systematic review", International Journal of Nursing Studies, vol. 102, 103466, 2020.

Semantically Query All (Squerall): A Scalable Framework to Analyze Data from Heterogeneous Sources at Different Levels of Granularity

Iqbal Hasan[1], Majid Zaman[2], Sheikh Amir Fayaz[3], Ifra Altaf [4], Muheet Ahmed Butt[5], S.A.M Rizvi[6]
Department of Computer Sciences, Jamia Millia Islamia, Delhi, India[1, 6]
Directorate of IT & SS, University of Kashmir, J&K, India[2]
Department of Computer Sciences, University of Kashmir, J&K, India[3, 4]
Directorate of IT & SS, University of Kashmir, J&K, India[5]

***Abstract*—In terms of data formats, codecs, and storage capacity, the previous two decades have seen a phenomenal progress. Rather than needing to adjust to one's application demands to the restricted storage solutions available in the past, there is now a wide range of options to pick from in order to best satisfy an application's needs. As a result, there is massive volume of data available in many forms and formats that, when linked and searched together, may yield significant knowledge and insights. We offer data warehouse as a solution and Big data as a solution in this study in order to handle heterogeneous data. However, both of these techniques have drawbacks when it comes to handling diverse data. Afterwards we propose another framework (Squerall), which relies on the ideas of Ontology-Based Data Access (OBDA) to allow querying of diverse heterogeneous sources using a single query language, SPARQL. In Squerall original data is queried on the fly, with no prior data materialization or modification. Squerall, in particular, enables the distributed aggregate and combining of massive data sets. It comes with five data sources out of the box, and it may be programmatically expanded to include other sources and query engines. The framework includes user interfaces for creating required inputs and for assisting non-SPARQL specialists through the process of writing SPARQL queries. It declares data mappings and transformations using RML and FnO, and employs Spark and Presto as query engines. The initiative underscored the importance of developing in this framework, technologies, and processes that enable for decentralized Big Data administration. Furthermore, demonstrating the feasibility and usefulness of OBDA on top of the growing NoSQL movement has a beneficial impact on Semantic Web principles acceptance. This demonstrates Squerall's importance and contribution to the organization.**

Keywords—Heterogeneous data; data warehouse; big data; presto; spark; squerall

I. Introduction

According to former Google CEO Eric Schmidt, "every two days today we produce as much content as we did from the birth of global civilization up until 2003, it's more like five Exabyte of data" [1]. In 2011, 300 million new websites were added, bringing the total number of websites to 555 million [2], resulting in a wide variety of data sources, each with its own structure and framework. User-desired quantitative approach remains an issue that must be understood and addressed as soon as possible. With the rapid rise of heterogeneous data, big data presents new challenges for extracting information from it. Scholars have acknowledged the usefulness of information that comes through interactions on social networks for many years. It is necessary to comprehend and use this information, as well as to evaluate the massive quantities that are increasing day by day [3]. These sources of information can contain structured, semi-structured, or unstructured data; that is, the records are in a variety of different formats such as relational model, HTML/XML file format, and so on, and different volumes raise key challenges in analytics on leveraging data features existing from multiple databases.

Higher education institutions, such as universities, also face similar challenges in dealing with massive amounts of data from a variety of sources for accreditation purposes, including data generated from online transactions, mobile phone applications, and data stored on multiple operating systems and repositories. This data is typically saved, categorized, retrieved, and evaluated in a standard style to fulfil accreditation criteria [4]. On the other hand, traditional database systems and technologies are incapable of effectively processing enormous volumes of data. Currently, developing software and methods for storing, managing, and manipulating enormous volumes of data is one of the most significant and hardest tasks in software systems research [5]. In addition, much of this information is kept in an unstructured way, utilizing a variety of languages and formats. Traditional techniques to data management are ineffective due to the data's vast quantity and complexity.

To deliver relevant insights, the analytical models require access to one or more data sources. Because of the differences in data type and amount, it is difficult to run any analytical model that can only function on a one data source. As a result, there is a need for a unique system that can execute analytic models using input from many data sources.

Heterogeneous data is unstructured information that is never presented in a structured fashion. Processing this unstructured data in a standard relational database is challenging. As a result, we favour a variety of techniques for dealing with diverse data. Data warehouse, big data, and Squerall-like systems are among these solutions. The former two, on the other hand, have their own set of repercussions when it comes to efficiently handling this diverse set of data

[6]. As a result, in this research, we use the Squerall architecture, which can handle unstructured data with billions of rows and millions of columns efficiently. Our research objective aims us to gain a better understanding of how diverse data is handled and how it varies with time.

The following section (Section II) deals with the backdrop of the current study, followed by Section III, which deals with the solution employing data warehouse and big data, as well as the limitations of each. A brief introduction to the Squerall algorithm is provided in Section IV, and the basic design and implementation are defined in Sections V and VI, respectively. The result analysis is covered in Section VII, as well as the overall operation of the approach employed in this study. Finally, Section VII wraps up the report with some recommendations for the future.

II. Literature Survey

Transforming and combining them into a single data source is a widely used approach for handling data heterogeneity issues [7-10]. When all of the data sources are combined into a single data source, the storage capacity need rises to that of a data warehouse, and some information may be lost as a result of lexical, syntactic, and geometrical discrepancies that occur during data integration.

S. A. Catanese [11] et al. described the collecting and analysis of massive data representing the linkages between online social network participants. They took two different approaches, both of which are well-defined and have been tested in practice against the popular social network Facebook. They presented two algorithms for data crawling: the BFS crawler and the Uniform crawler. For data integration, there are a variety of ways based on ontology. Bellatreche et al. [12] present the a priori approach, which is an ontology-driven data integration method. It is assumed that all data sources refer to a common ontology, which they may extend by adding their own concept specializations. There are two algorithms presented: 1: when sources add specific classes and properties to the shared ontology 2: when the shared ontology is extended by sources, but the occurrences are reflected into the shared ontologies before being exchanged.

Cruz et al. [13-14] presents a layered architecture for the integration of networked data that is syntactically, schematically, and semantically heterogeneous. They employ a global ontology to bridge the gap between the schemas of various data sources. A query is given in one of the data sources or the global ontology, and the mappings sharing of common lexicon are used to translate it into queries in other data sources.

Hashim [15] et al. present a cognitive integration framework that is integrated, mediated, and data warehouse-based. To guide the integration and handle syntactic, structural, and semantic heterogeneity, they use two types of ontologies: local and global ontologies.

Cur´e [16-17] et al. provides a mapping language to express access linkages to NoSQL databases for non-ontology-based access. It proposes an intermediate query language for converting SQL to Java methods for NoSQL database access. Query processing, on the other hand, is neither defined nor evaluated.

According to Gadepally [18] et al., data can be shifted across various databases at query time to improve calculation performance; the appropriate database is determined on a case-by-case basis. Although it has been demonstrated that using one database improves overall performance and data transportation, this has not been confirmed to be true with enormous data.

Support for a wide range of data sources is restricted or non-existent in all of the examined solutions. Only a few data sources [19-22] appear to be supported, and wrappers must be manually built or hard-coded. Squerall, on the other hand, does not invent the wheel and instead makes use of the numerous wrappers for existing motors. As a result, it is the solution with the most Big Data Variety support in terms of data sources.

III. Towards Solution

A. Data Warehouse as a Solution

"A data warehouse (DW) is a collection of data that has been arranged to be utilized as a decision support system," with data grouped by topics or subjects [23]. For instance, production, sales, and marketing. This arrangement enables for the collection of all relevant information on a given issue in order to aid decision-making. Users are less likely to modify or remove data in a warehouse since it is primarily used for consultation; this preserves the traceability of information so that analysis may be performed across time [24]. When data is put into a DW, data integration removes any conflicts of representation, resulting in a uniform representation of data.

The overall architecture of a DW is depicted in Fig. 1. It holds information from a variety of diverse and dispersed sources. Databases, data files, external sources, and other sources are examples of these sources. Data sources must be cleansed before being saved. The cleaning procedure involves selecting and purifying data in order to eliminate errors and resolve semantic disparities. The data will be integrated into the warehouse once it has been cleansed.

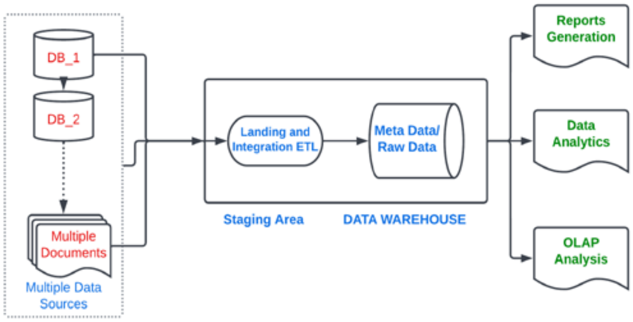


Fig. 1. Data Warehouse Architecture.

To integrate data from different sources, the ETL (extract, transform, and load) procedure must be employed, and data sources must be loaded at the logical warehouse schema transformations [25]. There are three steps to this: (1) Extraction is the process of retrieving data from several sources. To maintain data integrity, this stage necessitates synchronization of the extraction process. (2) Transformation is the process of formatting the extracted data using a set of rules to conform to the target warehouse schema. Assigning semantics to data sources and integrating source fields with target fields are two examples. (3) Loading is the process of transferring data from a source database, data warehouse, or data mart into a destination database, data warehouse, or data mart for analysis. A separate directory in the warehouse stores information on the warehouse's establishment, management, and use. This "metadata" information includes details on the schema sources, associated data, integration schema, refresh rules, user profiles, and user groups [26]. A data warehouse can be made up of many data marts. They are extracts from the warehouse that are tailored to a certain group of consumers and address a specific need. They work on OLAP (Online Analytical Processing) analysis and decision-making. OLAP delivers multidimensional representations of data to assist decision-making tools.

1) Limitations: Various data warehouses have been created in various sectors [26-27]. Today's DWs, on the other hand, confront fresh scientific obstacles. Indeed, today's data sources are diverse, self-contained, scalable, and dispersed. Traditional data warehouses face some limitations as a result of these challenges, which [28] summarizes as "Lack of scalability due to processing complexities coupled with inherent data issues and limitations of the underlying hardware, application software, and other infrastructure," and which we detail as follows:

- Data essence: New semi-structured and unstructured data models and formats have necessitated their integration and usage by modern data warehouses, yet classic DWs cannot manage semi-structured unstructured data [29].
- Data availability: The inability to obtain data at the proper time has an impact on the institution's decision-making use and implementation.
- Storage method and design: in the data warehouse, utilizing the same set of discs and controllers has a significant impact on both availability and performance analysis.

Because analytic and ad-hoc searches are non-deterministic, they have the largest influence on total query performance, data access, and processing, including data movement via the network [30]. Queries may call for a significant amount of data to be accessed from various storage regions, or a large amount of data to be accessed from a smaller storage area [31].

B. Big Data as a Solution

The fast creation of data by various technologies and people has reached enormous quantities that hardware and human capacities to interpret and modify cannot handle [32-33]. As a result, there is a common trend in the literature to conceive Big Data in terms of data volume. However, there is limited consensus on what defines Big Data in terms of volume among academics in many fields outside of education.

1) Limitations: Operating with Big Data systems necessitates the use of a high-speed computing infrastructure capable of handling vast amounts of data, which can be costly in terms of data acquisition, storage, analysis, and visualization [34]. Despite the fact that many academic institutions are already collecting numerous types of data, this information is stored in many databases, making analysis challenging. Furthermore, the absence of connectivity among institutional data systems makes gathering data from various systems for analysis difficult [35]. Additionally, the lack of data sharing agreements and data governance rules might act as a barrier to cross-institutional data integration and comparability [36]. Other key difficulties in Big Data systems include concerns about preserving personal and organizational privacy through verification and security.

Moreover, because much of the data already exists in institutional databases, current requirements for getting participants' consent in Big Data research are difficult to meet. Another ethical issue with utilizing Big Data for study is ensuring research integrity when using publically accessible data, because persons who created the data may not be willing to consent to its use, or such individuals may no longer be available for the study. Table I summarizes the full conclusion of the above-mentioned concerns.

Because the two viable solutions presented so far each have their own set of restrictions. Therefore, in addition to this, we need to discover a better solution that will address all of the difficulties that may occur in the future.

TABLE I. BIG DATA ISSUE SUMMARIZATION

Issues	Limitations
Management	Switching the entire company to a new infrastructure may be costly and time-consuming.
Storage	One Exabyte of data requires 25,000 gigabytes of disc space, which is difficult to manage and time-consuming to upload to the cloud.
Processing	Processing zettabytes and even Exabyte's of data appears to be a problem.
Heterogeneity	Big Data is frequently developed by combining many data sources belonging to various subpopulations. Each subpopulation may have some characteristics that are not shared by others. Data points from tiny subpopulations are often characterized as 'outliers' in traditional contexts when the sample size is small or moderate, and it is difficult to properly model them due to inadequate observations.
Noise Accumulation	When analyzing Big Data, we must concurrently estimate or test a large number of factors. When a decision or prediction rule is based on a large number of such factors, estimate mistakes accrue. In large dimensions, this noise accumulation effect is extremely severe, and it can even overwhelm real signals. The sparsity assumption is generally used to deal with it.

IV. Solution Beyond Data Warehouse and Big Data: Semantically Query all -- Squerall

Relational data management has been the dominant paradigm for storing and managing structured data for more than two decades. However, the introduction of really large-scale applications exposed relational data management's shortcomings in flexibly and horizontally extending the storing and querying of vast volumes of data. This sparked a paradigm change, requiring a new breed of databases capable of managing enormous data quantities while decreasing query expressivity and consistency constraints without compromising query efficiency.

By 2008, a slew of so-called non-relational or NoSQL (Not simply SQL) databases have popped up (e.g., MongoDB, Neo4j). One of the key Big Data difficulties is variety, which is exacerbated by heterogeneity. The fundamental rationale for the development of semantic technologies during the last two decades has been the integration of diverse data. Local data schemata are mapped to global ontology terms using defined mapping languages for a variety of popular data forms, such as relational data, JSON, CSV, or XML [37]. Heterogeneous data may therefore be accessed uniformly via queries written in SPARQL, a standardized query language that uses words from the ontology. Ontology-Based Data Access (OBDA) is the term used to describe this type of data access.

Three issues arise when implementing an OBDA architecture on top of Big Data:

- Translation of a query: SPARQL searches must be converted to the query languages of the various data sources. It's difficult to achieve a general and dynamic data model translation.
- Execution of federated queries: Because nonselective searches with big intermediate results are prevalent in Big Data settings, joining or aggregation cannot be executed on a single node but must be dispersed over a cluster.
- Information barriers: Data from a variety of sources can be linked to produce new insights, and it might not be easily 'joinable'.

Squerall, an extensible framework for querying Data Lakes, was built to address the above - mentioned issues [38].

- It enables ad hoc querying of massive, diverse data sources without the need for data integration or transformation.
- It enables the execution of distributed queries, particularly the combining of different heterogeneous sources.
- It allows users to describe query-time modifications that change join keys, allowing data to be joined.
- Squerall combines Apache Spark and Presto, two cutting-edge Big Data engines, with RML and FnO, two semantic technologies.

V. Squerall Architectural Framework

The OBDA principles [39] are used to build Squerall (Semantically query all). Most of these were designed to retrieve relational data but have no restrictions on the type or scale of data they may handle. We apply them to a Central Database, which contains massive and diverse data sources.

To help guide future debates, we will define the following concepts first:

- All notions used by data sources to characterize a single recorded datum, are represented by Data Attribute.
- A data entity, also known as a relevant entity, is a notion that data sources use to bring together related data, such as a table in a tabular database or a collection in a document database. There are one or more data characteristics associated with an object. If an entity includes information that matches a component of the query, it is relevant to the query.
- In response to a query, the Parallel Operational Area (POA) is the parallel distributed environment where ParSets are loaded, combined, and modified. It has an internal data structure that ParSets must follow.
- Any storage media, such as basic file storage or a database, is referred to as a data source.
- A data lake is a collection of many data sources where data is kept and retrieved in its original state and format, without being transformed beforehand [39].

A. Squerall Architectural Components

Squerall is made up of four primary parts (Fig. 2). We will utilize the generic ParSets and POA notions instead of Squerall's underlying corresponding concrete words, which vary from machine to machine, because of Squerall's extensible architecture.

1) Decomposer: Query Decomposer is a programme that decomposes queries. OBDA and query federation systems frequently use this component. The query's Basic Graph Pattern is decomposed into topic variables here. Because the focus of query execution is on bringing and merging entities from several sources, rather than retrieving a specific known item, query decomposition is subject-based (varying subjects). In a pre-processing phase, retrieving a certain entity, i.e., topic is constant, necessitates full-data parsing and index creation. This goes against the Data Lake notion of being able to access original data without having to go through a pre-processing stage. Nonetheless, a specific entity may be found by filtering on its properties.

2) Entity extractor: Relevant Entity Extractor is a programme that extracts relevant entities. This component searches the Mappings for entities with attribute mappings to each of the extracted star's characteristics for each extracted star.

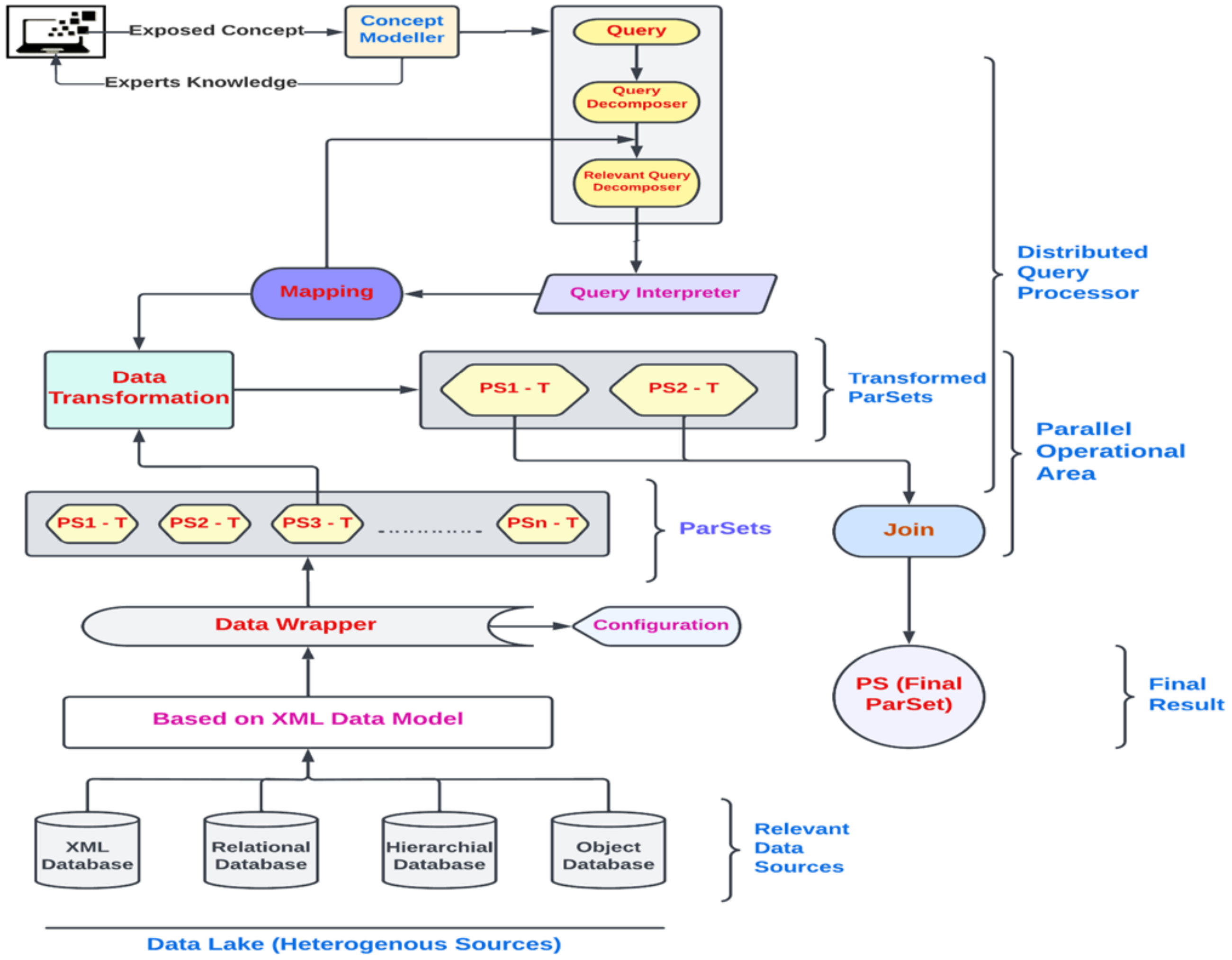


Fig. 2. Basic Squerall Architectural Components.

3) *Data wrapper:* The SPARQL query must be translated to the query language of the relevant data sources in the traditional OBDA. In the very varied Data Lake environments, this is difficult to implement in practice. As a result, a number of recent papers have recommended for the adoption of an intermediate query language. The intermediary query language in our example is POA's query language, which is determined by its internal data structure. At query time, the Data Wrapper creates data in POA's data structure, allowing for the concurrent execution of costly operations such as join. Wrappers must exist to transform data entities from the source to POA's data structure, either completely or partly, depending on whether parts of the data may be sent down to the original source.

4) *Distributed query processor:* The final results are formed by joining ParSets together. In the POA, ParSets may be subjected to any query action, including as selection, aggregation, and ordering. However, because we are interested in querying several data sources, we will concentrate on the join process.

VI. Squerall Implementation

Scala is used to write Squerall. It declares data mappings and transformations using RML and FnO, and employs Spark and Presto as query engines.

A. Mapping

Squerall allows RML entity and attribute mappings, which extends the W3C R2RML mapping language to facilitate mapping of diverse sources.

The rml:logicalsource fragment, which specifies the entity source and type, is anticipated.

rr:subjectMap is used to obtain the entity ID.

rr:predicateObjectMap, which is used for all entity attributes and translates an rml:reference to an ontology word through rr:predicate.

We extend RML with the NoSQL: store attribute from our NoSQL ontology, which allows you to define the entity type, such as Cassandra, MongoDB, and so on.

B. Information Transformation

Users can specify transformations in Squerall to provide data join ability. There are two conditions that must be met:

1) The transformation specification must be divorced from the technical implementation.

2) Transformations must be done on-the-fly at query time, as defined by the Data Lake definition.

We employ the Function Ontology, which abstracts away the particular technology used to describe machine-processable high-level functions [40-42]. We utilize *FnO* in combination with RML in the same way as the *DBpedia* Extraction Framework does. We do not, however, build RDF triples physically; instead, we apply *FnO* operations on-the-fly at query time.

When a transformation declaration is satisfied, Squerall examines the mappings at query time and initiates appropriate Spark and Presto actions over the query intermediate results [43]. A map () transformation is employed in Spark, whereas equivalent string or number SQL operations are used in Presto.

C. Querying and Wrapping

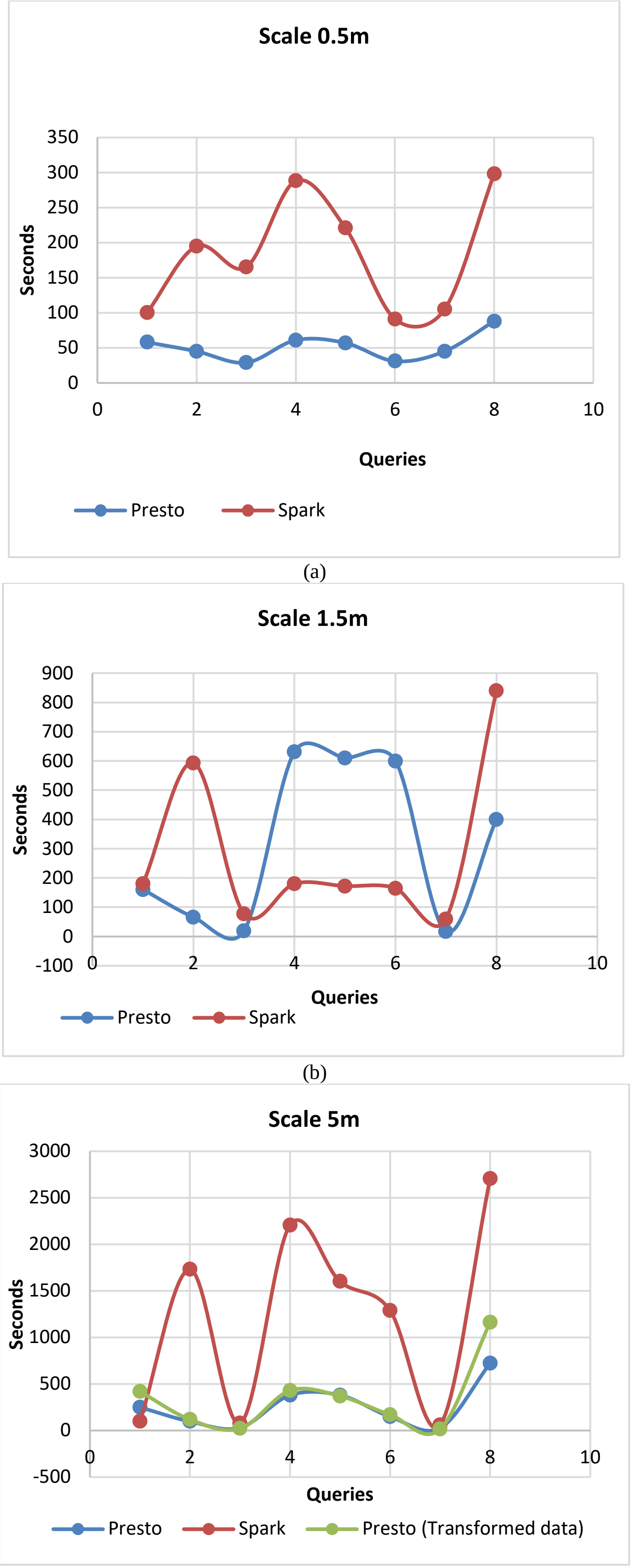
We use two popular frameworks to build the Squerall engine: Apache Spark and Presto. Spark is a general-purpose processing engine, while Presto is a distributed SQL query engine for active querying. Both rely on memory for their operations [44-45]. We use the connector idea in Spark and Presto, which is a wrapper that can import data from an external source into their internal data structure (*ParSet*), flattening any semi representations [46]. DataFrame is Spark's internal data structure, which is a tabular format that can be queried programmatically in SQL. Their structure is the same as the *ParSet's*, with one column per star predicate.

VII. Result Analysis and Discussion

We assess the correctness of the results as well as the query's performance. We compare query results to a centralized relational database (MySQL) for correctness, since it reflects data at its most consistent level. We test query execution time against the three created scales to determine performance. The influence of the number of joins on query time is highlighted in detail. We determine the mean value of each query after running it three times. The timeout value is set to 3600 seconds.

In the absence of a similar study that allows to query all five data sources and SPARQL fragments that Squerall supports, we compare the performance of Squerall's two underlying query engines: Spark and Presto.

In case of Accuracy measure, Squerall provided the same amount of returns in all scale 0.5m queries as MySQL, indicating 100 percent accuracy. We evaluated the effectiveness of the various engines after MySQL timed out with data of scale 1.5m, and the results were likewise similar.



(a)

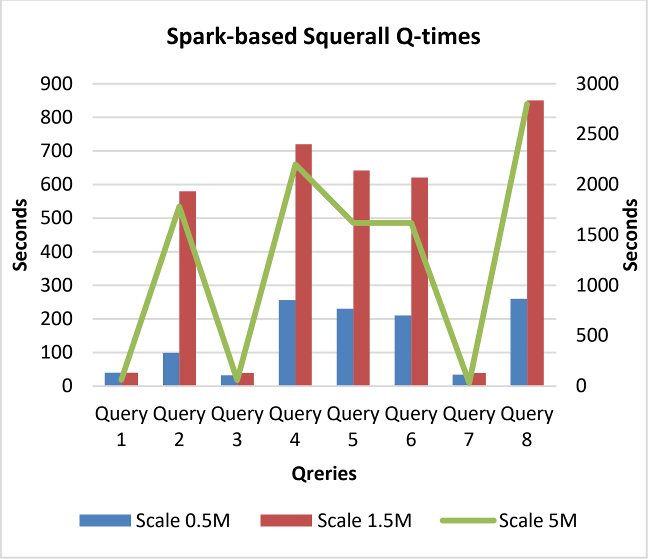
(b)

(c)

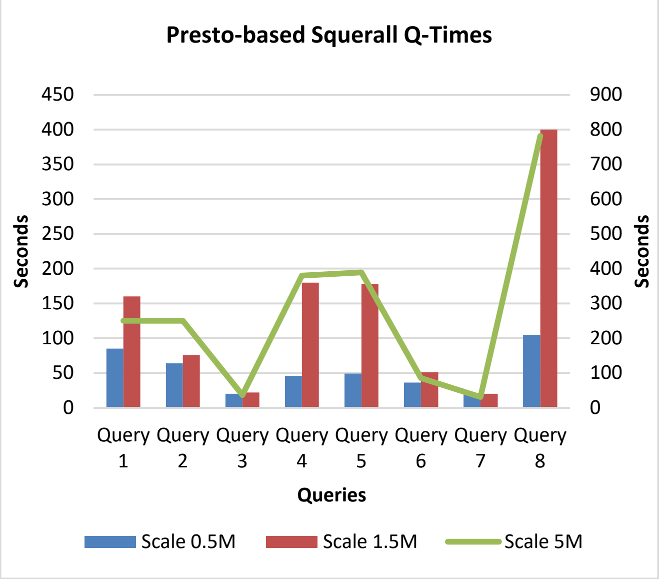
Fig. 3. Query Execution Time based on Presto, Spark on different Scales (0.5m, 1.5m & 5m).

While as, the findings imply that Squerall performs rather well across a wide range of queries, i.e., a large number of them. Filtering and sorting of joins, with and without filtering Squerall, located in Presto, is on display. Up to an order of magnitude greater performance than Spark-based solutions. Query performance is higher across the board with the 0.5m [Figure 3(a)] data scale. There was a rise of up to 800% in the number of inquiries. Presto-based models come in 1.5m [Figure 3(b)] and 5m scales [Figure 3(c)]. With a rise of up to 1300 percent in all inquiries except Query 1, is outstanding. Thus, a variety of elements contributes to greatness.

Query performance did not suffer as data size increased; query times were roughly proportional to data size (Fig. 4) and stayed below the threshold. We purposely add modifications to the data so that it becomes unjoinable in order to test the effect of the query-time data manipulations.



(a)



(b)

Fig. 4. Percentage differences between the Scales (Presto-based and Spark-based).

The findings reveal that in the vast majority of cases, the cost is insignificant. This is due to the fact that both Spark and Presto use RAM to do computations. These transformations in Spark (figure 4a) only use the map function, which is done locally and does not require any data movement. Only a few queries out of 5 million in Presto-based Squerall [Figure 4(b)] had expenses that were evident but not significant. We only add the results of the processing cost at the scale 5M due to small variances and to improve readability. Our findings might be viewed as a comparison of results between Spark and Presto, which is a rare occurrence.

VIII. Conclusion and Future Strategies

We described Squerall in this paper as a framework for achieving the Semantic Data Lake, or querying diverse and huge data sources using Semantic Web approaches. It lets users to describe changes that enable joinability on-the-fly at query time, and it conducts distributed cross-source join operations. Squerall is designed with Spark and Presto, two cutting-edge Big Data technologies. Squerall removes users from handcrafting wrappers by relying on the latter's connectors to wrap the data—a fundamental bottleneck in supporting data variation across the literature. It also allows Squerall to be readily extended. Squerall may also be dynamically modified to use different query engines thanks to its modular code design. We want to enable more SPARQL operations in the future, such as OPTIONAL and UNION, as well as take use of the query engines' own algorithms to improve query performance. Finally, there is a natural necessity to maintain provenance at the data and query results levels in such a diverse setting.

References

[1] Zaman, M. and Butt, M.A., 2012, October. Information translation: a practitioners approach. In *World Congress on Engineering and Computer Science (WCECS).*

[2] Fayaz, Sheikh Amir, Ifra Altaf, Aaqib Nazir Khan, and Zahid Hussain Wani. "A possible solution to grid security issue using authentication: an overview." J. Web Eng. Technol 5, no. 3 (2019): 10-14.

[3] Zaki, R., Barabadi, A., Barabady, J. and Nouri Qarahasanlou, A., 2022. Observed and unobserved heterogeneity in failure data analysis. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *236*(1), pp.194-207.

[4] Bryan, Christopher J., Elizabeth Tipton, and David S. Yeager. "Behavioural science is unlikely to change the world without a heterogeneity revolution." *Nature human behaviour* 5, no. 8 (2021): 980-989.

[5] Matthewes, S.H., 2021. Better together? Heterogeneous effects of tracking on student achievement. *The Economic Journal*, *131*(635), pp.1269-1307.

[6] Fathy, Naglaa, Walaa Gad, Nagwa Badr, and Mohamed Hashem. "ONTOLOGY-BASED DATA ACCESS TO HETEROGENEOUS DATA SOURCES: STATE OF THE ART APPROACHES AND APPLICATIONS." *International Journal of Intelligent Computing and Information Sciences* (2022): 1-10.

[7] Fayaz, Sheikh Amir, Majid Zaman, and Muheet Ahmed Butt. "Numerical and Experimental Investigation of Meteorological Data Using Adaptive Linear M5 Model Tree for the Prediction of Rainfall." *Review of Computer Engineering Research* 9, no. 1 (2022): 1-12.

[8] Kaul, S., Fayaz, S.A., Zaman, M. and Butt, M.A., 2022. Is decision tree obsolete in its original form? A burning debate. *Revue d'Intelligence Artificielle*, *36*(1), pp.105-113.

[9] Fayaz, Sheikh Amir, Majid Zaman, and Muheet Ahmed Butt. "Knowledge Discovery in Geographical Sciences—A Systematic Survey of Various Machine Learning Algorithms for Rainfall

Prediction." In *International Conference on Innovative Computing and Communications*, pp. 593-608. Springer, Singapore, 2022.

[10] Fayaz, Sheikh Amir, Majid Zaman, and Muheet Ahmed Butt. "An application of logistic model tree (LMT) algorithm to ameliorate Prediction accuracy of meteorological data." International Journal of Advanced Technology and Engineering Exploration 8, no. 84 (2021): 1424-1440.

[11] S. Catanese, P. De Meo, E. Ferrara, G. Fiumara, A. Provetti: Crawling Facebook for Social Network Analysis Purposes,WIMS '11 Proceedings of the International Conference on Web Intelligence, Mining and Semantics, 2011, Article no: 52.

[12] Ladjel Bellatreche, Dung Nguyen Xuan, Guy Pierra, and Hondjack Dehainsala. 2006. Contribution of ontology-based data modeling to automatic integration of electronic catalogues within engineering databases. Computers in Industry 57, 8-9 (2006), 711–724.

[13] Isabel F. Cruz and Huiyong Xiao. 2005. .e role of ontologies in data integration. Engineering Intelligent Systems 13, 4 (2005), 245–252.

[14] Isabel F. Cruz and Huiyong Xiao. 2009. Ontology Driven Data Integration in Heterogeneous Networks. In Complex Systems in Knowledge-based Environments: Theory, Models and Applications. 75–98.

[15] H.A.Hashim, A. Ahmed, N. Salim, A.Osman O.Y.Sheng, A.Sim, A.Bakri, N.H.Zakaria, R.Ibrahim, and S.S.Omar. 2005. A New Database Integration Model Using An Ontology-Driven Mediated Warehousing Approach. Journal of Theoretical and Applied Information Technology 58, 2 (2005), 392–409.

[16] Cur´e, O., Kerdjoudj, F., Faye, D., Le Duc, C., Lamolle, M.: On the potential integration of an ontology-based data access approach in NoSQL stores. Int. J. Distrib. Syst. Technol. (IJDST) 4(3), 17–30 (2013).

[17] Cur´e, O., Hecht, R., Le Duc, C., Lamolle, M.: Data integration over NoSQL stores using access path based mappings. In: Hameurlain, A., Liddle, S.W., Schewe, K.-D., Zhou, X. (eds.) DEXA 2011. LNCS, vol. 6860, pp. 481–495. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23088-2 36.

[18] Gadepally, V., et al.: The BigDAWG polystore system and architecture. In: High Performance Extreme Computing Conference, pp. 1–6. IEEE (2016).

[19] Fayaz, Sheikh Amir, Majid Zaman, and Muheet Ahmed Butt. "To ameliorate classification accuracy using ensemble distributed decision tree (DDT) vote approach: An empirical discourse of geographical data mining." Procedia Computer Science 184 (2021): 935-940.

[20] Fayaz, Sheikh Amir, Majid Zaman, Sameer Kaul, and Muheet Ahmed Butt. "Is Deep Learning on Tabular Data Enough? An Assessment."

[21] Fayaz, Sheikh Amir, Majid Zaman, and Muheet Ahmed Butt. "Performance Evaluation of GINI Index and Information Gain Criteria on Geographical Data: An Empirical Study Based on JAVA and Python." In International Conference on Innovative Computing and Communications, pp. 249-265. Springer, Singapore, 2022.

[22] Altaf, Ifra, Muheet Ahmed Butt, and Majid Zaman. "A Pragmatic Comparison of Supervised Machine Learning Classifiers for Disease Diagnosis." In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1515-1520. IEEE, 2021.

[23] Mutanu, L., & Khan, Z. (2021). Enhancing the QoS of a Data Warehouse through an Improved ETL Approach. Journal of Language, Technology & Entrepreneurship in Africa, 12(2), 89-102.

[24] Kormeier, B., & Hippe, K. (2022). Data Warehousing of Life Science Data. Integrative Bioinformatics, 85-96.

[25] Golfarelli, Matteo, and Stefano Rizzi. Data warehouse design: Modern principles and methodologies. McGraw-Hill, Inc., 2009.

[26] Sebaa, A., Chikh, F., Nouicer, A., & Tari, A. (2017). Research in big data warehousing using Hadoop. Journal of Information Systems Engineering & Management, 2(2), 10.

[27] Wu, L., Sumbaly, R., Riccomini, C., Koo, G., Kim, H. J., Kreps, J., & Shah, S. (2012). Avatara: OLAP for web-scale analytics products. Proceedings of the VLDB Endowment, 5(12), 1874-1877.

[28] Krishnan, K. (2013). Data warehousing in the age of big data. Newnes.

[29] Mohd, R., Butt, M. A., & Baba, M. Z. (2020). GWLM–NARX: Grey Wolf Levenberg–Marquardt-based neural network for rainfall prediction. *Data Technologies and Applications*.

[30] Fayaz, S.A., Kaul, S., Zaman, M., Butt, M.A. (2022). An adaptive gradient boosting model for the prediction of rainfall using ID3 as a base estimator. Revue d'Intelligence Artificielle, Vol. 36, No. 2, pp. 241-250. https://doi.org/10.18280/ria.360208.

[31] Fayaz SA, Zaman M, Butt MA. A hybrid adaptive grey wolf Levenberg-Marquardt (GWLM) and nonlinear autoregressive with exogenous input (NARX) neural network model for the prediction of rainfall. International Journal of Advanced Technology and Engineering Exploration. 2022; 9(89):509-522. DOI:10.19101/IJATEE.2021.874647.

[32] Vaitsis, Christos, Vasilis Hervatis, and Nabil Zary. "Introduction to big data in education and its contribution to the quality improvement processes." Big Data on Real-World Applications 113 (2016): 58.

[33] Vaitsis, Christos, Vasilis Hervatis, and Nabil Zary. "Big Data on Real-World Applications. Chapter 3: Introduction to Big Data in Education and Its Contribution to the Quality Improvement Processes." (2016).

[34] Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information sciences, 275, 314-347.

[35] Rathore, M. Mazhar, Awais Ahmad, Anand Paul, and Alfred Daniel. "Hadoop based real-time big data architecture for remote sensing earth observatory system." In 2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-7. IEEE, 2015.

[36] Miyares, J., & Catalano, D. (2016). Institutional analytics is hard work: a five-year journey. EDUCAUSE review September/October 2016, pp. 8–9. Retrieved September 1, 2016, from http://er.educause.edu/~/media/files/articles/2016/8/erm1656.pdf.

[37] Michel, F., Faron-Zucker, C., Montagnat, J.: A mapping-based method to query MongoDB documents with SPARQL. In: Hartmann, S., Ma, H. (eds.) DEXA 2016. LNCS, vol. 9828, pp. 52–67. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44406-2-6.

[38] Mami, M.N., Graux, D., Scerri, S., Jabeen, H., Auer, S.: Querying data lakes using spark and presto (2019, To appear in The WebConf - Demonstrations).

[39] Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R.: Linking data to ontologies. In: Spaccapietra, S. (ed.) Journal on Data Semantics X. LNCS, vol. 4900, pp. 133–173. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-77688-8-5.

[40] Vogt, M., Stiemer, A., Schuldt, H.: Icarus: towards a multistore database system. In: 2017 IEEE International Conference on Big Data (Big Data), pp. 2490–2499 (2017).

[41] Altaf, Ifra, Muheet Ahmed Butt, and Majid Zaman. "Disease Detection and Prediction Using the Liver Function Test Data: A Review of Machine Learning Algorithms." In International Conference on Innovative Computing and Communications, pp. 785-800. Springer, Singapore, 2022.

[42] Ashraf, M., Salal, Y. K., Abdullaev, S. M., Zaman, M., & Bhut, M. A. (2022). Introduction of Feature Selection and Leading-Edge Technologies Viz. TENSORFLOW, PYTORCH, and KERAS: An Empirical Study to Improve Prediction Accuracy of Cardiovascular Disease. In International Conference on Innovative Computing and Communications (pp. 19-31). Springer, Singapore.

[43] Hassan, Musavir, Muheet Ahmed Butt, and Majid Zaman. "An Ensemble Random Forest Algorithm for Privacy Preserving Distributed Medical Data Mining." International Journal of E-Health and Medical Communications (IJEHMC) 12, no. 6 (2021): 1-23.

[44] Shehloo, Arif Ahmad, Muheet Ahmed Butt, and Majid Zaman. "Factors affecting cloud data-center efficiency: a scheduling algorithm-based analysis." International Journal of Advanced Technology and Engineering Exploration 8, no. 82 (2021): 1136.

[45] S. J. Sidiq, M. Zaman, and M. Ahmed, "How machine learning is redefining geographical science: A review of literature," International Journal of Emerging Technologies and Innovative Research, vol. 6,, pp. 1731-1746, 2019.

[46] Fayaz, S.A., Zaman, M., Kaul, S., Butt, M.A. (2022). How M5 model trees (M5-MT) on continuous data are used in rainfall prediction: An experimental evaluation. Revue d'Intelligence Artificielle, Vol. 36, No. 3, pp. 409-415. https://doi.org/10.18280/ria.360308.

Evaluation on the Effects of 2D Animation as a Verbal Apraxia Therapy for Children with Verbal Apraxia of Speech

Muhammad Taufik Hidayat[1], Sarni Suhaila Rahim[2]*, Shahril Parumo[3]
Nurul Najihah A'bas[4], Muhammad 'Ammar Muhammad Sani[5], Hilmi Abdul Aziz[6]
Fakulti Teknologi Maklumat Dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM), 76100 Melaka, Malaysia[1, 2, 3, 4, 5]
EDUfa Autism Therapy Centre Jakarta Timur, Jl. Jambore No.11, RT.5/RW.11, Cibubur, Kec. Ciracas, Kota Jakarta Timur[6]
Daerah Khusus Ibukota Jakarta 13720, Indonesia[6]

***Abstract*—This article presents an evaluation of 2D Animation of Learning Numbers and Letters for Children with Verbal Apraxia. The developed application provides some knowledge and encourage children with verbal apraxia to learn and know about numbers and letters. An experimental testing was conducted to evaluate the usability of the developed application, aimed as a therapy for the children who suffer this apraxia across all age. Five important evaluation components such as learnability, usability, accessibility, functionality, and effectiveness were included in this testing to investigate the user engagement and satisfaction of the proposed medical and educational learning system. Online questionnaires were distributed as a method to collect user testing outputs. A total of 33 respondents from multimedia designers, practitioners, psychologists, and parents were involved in this survey. The results of the testing indicate that majority of respondents are satisfied with the outcomes of the 2D animation video. The results presented may facilitate improvements in the teaching syllabus for students with speech and language disorder and produce a great visual animation treatment to the users.**

***Keywords*—Childhood apraxia of speech; verbal dyspraxia; speech and language disorder; 2D animation; visual animation treatment; evaluation**

I. Introduction

Childhood Apraxia of Speech (CAS) is a speech condition in which a child's brain struggles to coordinate the complex oral motions required to create sounds into syllables, syllables into words, and words into phrases. An evaluation of the child's expressive and receptive language abilities is required as part of a diagnosis of childhood apraxia of speech; many children with this disorder have language deficiencies. This article focuses on the testing phase of the 2D animation of Learning Numbers and Letters for Children with Verbal Apraxia. The objective of this paper is to present a comprehensive usability evaluation of visual animation named "2D animation of Learning Numbers and Letters for Children with Verbal Apraxia". There are five evaluation components included in the testing to evaluate the user engagement and satisfaction of the 2D animation learning system. It is envisaged the proposed product would assist in providing a visual therapy to patients with Apraxia of Speech in the form of 2D animated video. The contribution of this study is obvious as the resulting outcomes can be capitalised as guidelines to increase an awareness about Verbal Apraxia to the public.

II. Literature Review / Previous Work

There have been several studies in the literature reporting on the usage of animation, particularly in teaching and learning purpose [1-3]. In addition, there are some products which are related to the proposed visual animation for verbal apraxia of speech. For example, Speech Therapy Songs Animation by Barefoot Books [4] and Alphabets and Numbers Verbal Therapy by Terapi Autisme Online [5]. However, these products are limited for general, autism and spectrum disorder use. Besides, these applications are not interactive and the syllabus is outdated.

Hence, a 2D Animation of Learning Numbers and Letters for Children with Verbal Apraxia [6] is developed. This application is focusing on visual therapy to patients with Apraxia of Speech using an animation element. The comprehensive explanation on the design and development phases of the application; 2D Animation of Learning Numbers and Letters for Children with Verbal Apraxia is presented in [6]. The comparison of the existing products with the proposed application is also presented in [6]. Fig. 1, Fig. 2 and Fig. 3 present the screenshots of the developed 2D animation of Learning Numbers and Letters.

Fig. 1. Screenshot of the Main Page.

*Corresponding Author.

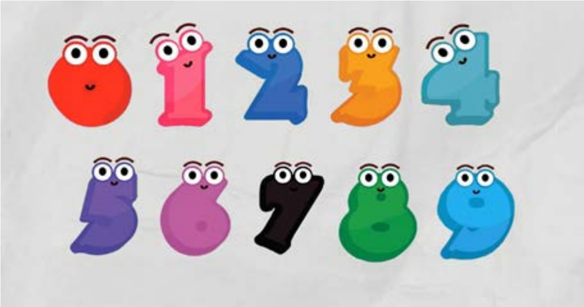


Fig. 2. Screenshot of the Number Module Page.

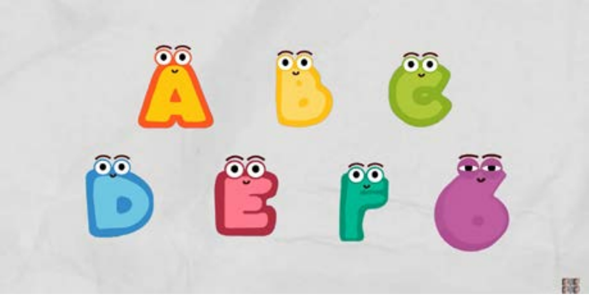


Fig. 3. Screenshot of the Alphabet Module Page.

Usability is described as a product's or service's ability to provide maximum satisfaction, efficiency, and effectiveness when utilised by various types of users as described in [7]. The usability has been investigated by many researchers in [8-13]. In recent years, there has been an increasing amount of literature on usability testing [14-15].

This paper presents an extension of the work reported earlier in [6] with emphasis on the evaluation of the proposed application. Also, this paper elaborated five components which are used as evaluation components for the system testing purpose; they are learnability, usability, accessibility, functionality, and effectiveness.

III. Methodology

The methodology section presents about the testing and outcomes. The test plan comprising of the test user, test schedule, test strategy and test implementation and eventually test results are elaborated in this section.

In this research work, the main target user is children who suffers from Verbal Apraxia of Speech from the age of 5-10 in Autism Therapy Center in Edufa Pekanbaru, Indonesia. The practitioner who guides the children will also be the target user as they will be observing the children’s behavior while checking the content of the application to make sure it is according to the children’s method of study. Table I shows the details of respondents that involved in this testing.

1) Multimedia expert: A multimedia expert is a specialist in the field of multimedia and information technology who has extensive knowledge and expertise. Graphic Designer, Animator, Videographer, and Lecturer were chosen to take the test. This test is performed at the end of the development process and before the product is released. They will test the application with a focus on interface, interactivity, design, integration of multimedia elements, and content layouts in the application.

2) Subject matter expert: A subject matter expert is an individual that has a specialties and background of medical that help improve the application before and after the testing method. The practitioners, therapists, and psychologists were chosen to take the test. They will independently test the application and give an input about the application with medical terms and medical point of view.

3) Parents: The children with autistic and verbal apraxia of speech are the main target user of this research work. Since the testing was conducted online, the device setup and supervision were assisted by their parents. This category is addressed to the parents, who will explain the questionnaire to their children with their own way. Basically, this category is answered by their children assisted by the role of parents.

During the testing phase, test implementation will outline how the testing will be implemented to a certain target user. The relationship between the test description and the test data is carried out in accordance with the test strategy.

A. Test Description

The total number of respondents for the testing procedure is 33 people. After receiving an explanation of the research work, each respondent will conduct the testing independently. They are required to answer all of the questionnaire questions. Each of the respondent is required to respond and provide comments regarding the application in their perspective.

B. Test Data

The test data for the user testing will be explained in Table II, while Table III, Table IV and Table V represent the respondents of subject matter expert, multimedia expert and parents, respectively.

TABLE I. TESTING RESPONDENTS

	Multimedia Expert	**Subject Matter Expert**	**Parents**
General Information	Practitioners who teach in Autism Center, and Doctor who expert in Autism	Multimedia Designer who works and experienced in multimedia field	Parents of children with Apraxia of Speech
Description	To evaluate the usage and the effectiveness of multimedia elements in the animation	To identify the content validity, provides understanding related to subtopics	To identify the effectiveness of the animation as a therapy

TABLE II. TEST DATA FOR USER TESTING

General Information	**Number of Respondents**
Practitioners who teach in Autism Center, and Doctor who expert in Autism	16
Multimedia Designer who works and experienced in multimedia field	11
Parents with autistic children, especially for children with verbal apraxia of speech	6

TABLE III. DETAILS OF SUBJECT MATTER EXPERT

Respondents	Gender	Age	Occupation
Respondent 1	Female	28	Practitioner at Palembang Therapy Center
Respondent 2	Male	26	General Practitioner at Eka Hospital Pekanbaru
Respondent 3	Female	34	Lecturer at Sriwijaya University
Respondent 4	Female	29	Teacher at Palembang School with Special Needs
Respondent 5	Female	23	Assistant Psychologist at Magna Penta
Respondent 6	Female	25	General Practitioner at Palembang Therapy Center
Respondent 7	Female	27	Therapist at Clemira Therapy Center
Respondent 8	Female	25	Therapist at Clemira Therapy Center
Respondent 9	Female	23	Psychology Student, Intern at Palembang Health Department
Respondent 10	Female	23	Psychology Student, Volunteer at Palembang Autism Parenting 2020
Respondent 11	Female	32	Psychologist, Psychotherapy at Optime Palembang
Respondent 12	Male	22	Psychology Student, Volunteer at Hari Peduli Autisme
Respondent 13	Male	28	Child Therapist at Dian Selaras Konsultan Psikologi & Klinik Hipnoterapi
Respondent 14	Female	22	Psychology Student at Sriwijaya University
Respondent 15	Female	23	Psychology Student at Sriwijaya University
Respondent 16	Female	22	Psychology Student at Sriwijaya University

TABLE IV. DETAILS OF MULTIMEDIA EXPERT

Respondents	Gender	Age	Occupation
Respondent 1	Male	24	Graphic Designer at RAGBI MALAYA RESOURCES
Respondent 2	Male	21	Animator at KINARYA COOP, Film and Animation Student at Binus University
Respondent 3	Male	21	Freelance Videographer at Tanjak Studio
Respondent 4	Male	21	Freelance Filmmaker, Broadcasting student at Ahmad Dahlan University
Respondent 5	Male	22	Video Editor Intern at Sinemalis Studio
Respondent 6	Male	22	Deputy Coordinator of Multimedia Bureau at PPI Malaysia
Respondent 7	Female	24	Web Developer Intern at Sma Maahad As-Syakhsiah Tahfiz Sains, Pulau Pinang
Respondent 8	Male	23	Project Manager at PT. Kreatif Media Industri
Respondent 9	Female	24	Lecturer at University of Mataram
Respondent 10	Female	24	Master of Computer Science at Universiti Teknikal Malaysia Melaka
Respondent 11	Female	25	Graphic Designer at DigitalBrain SDN BHD

TABLE V. DETAILS OF PARENTS

Respondents	Gender	Age
Respondent 1	Female	43
Respondent 2	Female	42
Respondent 3	Male	36
Respondent 4	Male	41
Respondent 5	Female	32
Respondent 6	Male	40

C. *Chart Results and Analysis*

Diagrams and charts will be displayed in this analysis based on the results of the overview and testing measures. This is an examination diagram of the assessment testing that was conducted. A few charts have been created based on the information obtained from the testing results to summarize the outcome of the assessment.

1) Multimedia expert: Eleven multimedia experts, including animators, graphic designers, videographers, filmmakers, lecturers, and multimedia students, took part in the testing phase. The multimedia experts were asked to assess the usage of the multimedia features in the animation video, which include content, audio, video, and interface design. The collected data will be analyzed and compiled into a bar chart.

a) Chart of Learnability for Multimedia Expert: Fig. 4 shows the testing results from multimedia expert for the learnability aspect. For the information and messages outcome, we can see that 63.6% of multimedia experts strongly agree that the information and message in animation are simple and straightforward, this is because animation is aimed at children, who do not like watching long videos that make them feel bored.

There are about 72.7% of multimedia experts who strongly agree that the content of the animation is very easy to understand because the animation displays eye-catching and easy-to-understand visuals for the viewers.

It is indicated that there are 54.5% experts agreed that the instructions in the video can guide users because the instructions are easy to follow and understandable, but parental assistance is needed to be able to help children to always follow the instructions until the end.

b) Chart of Usability for Multimedia Expert: Fig. 5 shows the testing results from multimedia expert for the usability aspect. Although 45.5% agree, 9.1% disagree that the text and graphics are readable and clear because some experts feel that these graphics need to be improved, so that they are easily seen by the viewer's eyes.

More than 50% of experts agree that users do not need help with instructions outside of the video because the instructions in the video itself are very clear to understand and follow.

As illustrated in Fig. 5, the finding revealed that 54.5% of experts agree that the use of animation gains a better understanding of the content of the animated video because of the interaction of the host and the movement of the animation to the viewer.

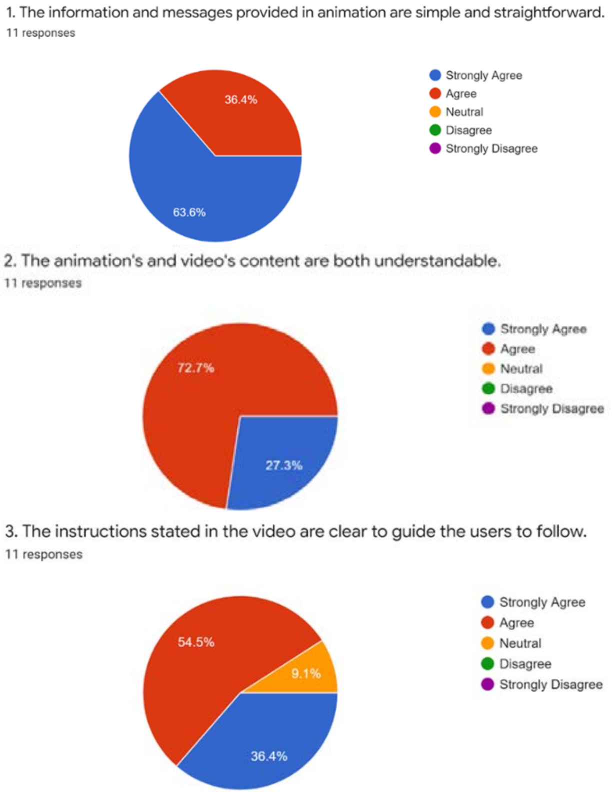


Fig. 4. Result of Learnability by the Multimedia Experts.

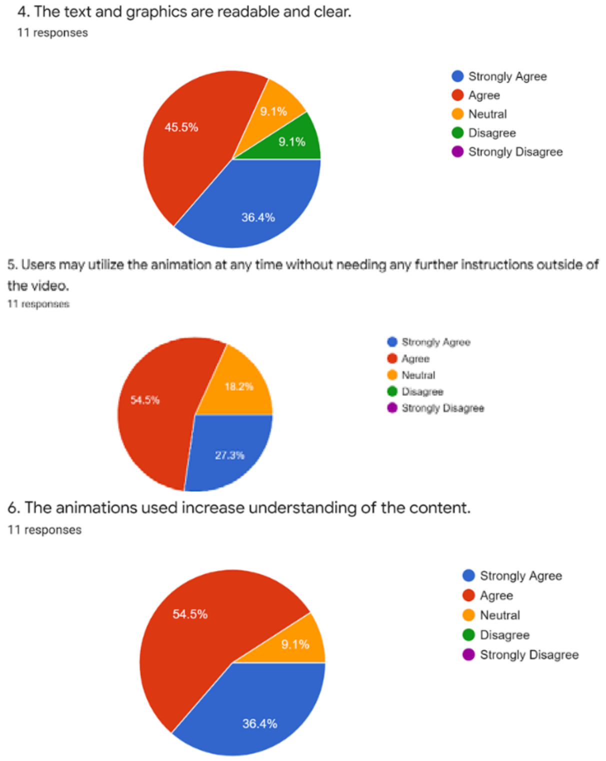


Fig. 5. Result of Usability by the Multimedia Experts.

c) Chart of Effectiveness for Multimedia Expert: Fig. 6 shows the testing results from multimedia expert for the effectiveness aspect. Fig. 6 shows more than half experts agree that integration between multimedia elements can help users obtain information effectively because multimedia instruments work together to help deliver information.

About 72.7% of experts agree that the layout of the content increases the effectiveness in conveying information as illustrated in Fig. 6 because the arrangement in the preparation of the layout is considered by the creators before being applied to the video.

Findings in Fig. 6 highlight 72.7% of experts agree that the information has an impact on the user as stated in Fig. 12 because they believe that the instructions and interactions will help children in verbal apraxia therapy through visual interactions.

d) Chart of Accessibility for Multimedia Expert: Fig. 7 shows the testing results from multimedia expert for the accessibility aspect. Fig. 7 clarifies that although seven respondents agree that audio used is suitable, it turns out that there is one respondent who thinks that the audio used is not clear to hear because indeed the device in using audio recording does not use a special microphone for voice over, but only an ordinary cellphone headset.

Surprisingly two experts believe that the graphics used are not too attractive. This is because there is still a lot of potential that can be developed from the animation, such as designs that are more attractive to children than the existing ones.

Fig. 7 shows that almost as many as a quarter of experts believe that the use of color in this animation is appropriate and attractive because the purpose of this animation is for children, so the animation must be made as interesting as possible.

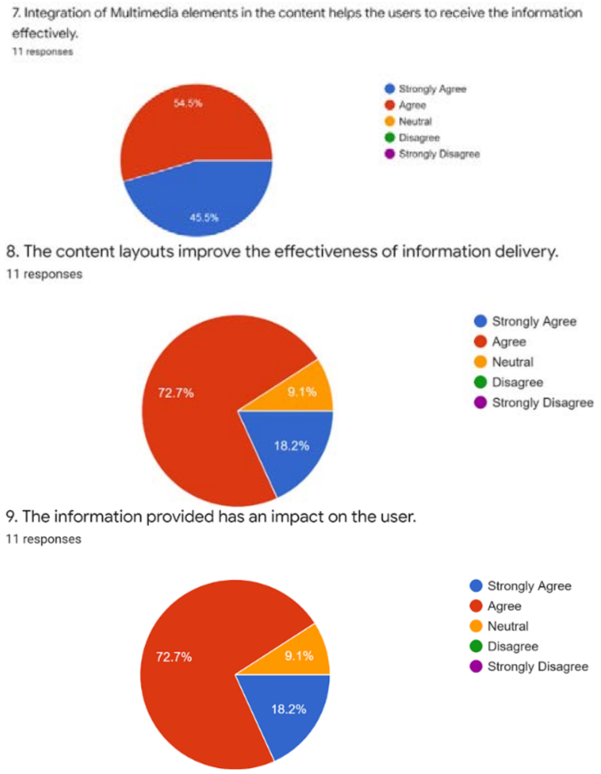


Fig. 6. Result of Effectivess by the Multimedia Experts.

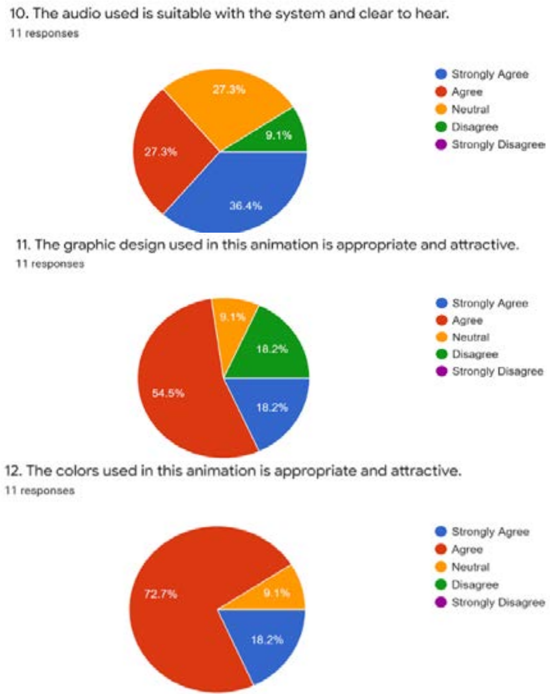


Fig. 7. Result of Accessibility by the Multimedia Experts.

e) Chart of Functionality for Multimedia Expert: Fig. 8 shows the testing results from multimedia expert for the accessibility aspect. Fig. 8 illustrates that there are four experts who strongly agree, and four others also agree with the statement that the animation is smooth and can be seen clearly. This is because according to their proposal, the use of visual aspects is very critical in making an animation video.

This study indicates that 63.6% experts agree that the user will be able to catch the message from the host quickly as presents in Fig. 8, this is because of the influence of good delivery and interaction from the host which makes the message will be conveyed well to the users.

Summary for Multimedia Expert responses are summarises in Table VI.

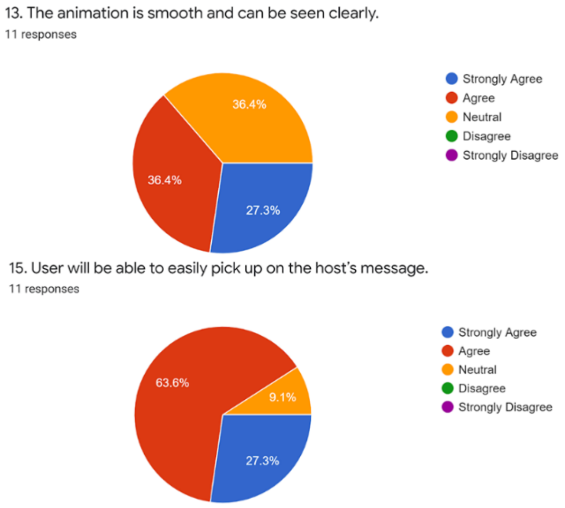


Fig. 8. Result of Functionality by the Multimedia Experts.

f) Suggestion for Improvement from Multimedia Expert: Based on the questionnaire to the multimedia experts, below are the suggestion for improvement:

i. Font for animation (numbers and alphabets) can be improved if the font used is a more familiar font. For example, the number 0 looks like a dot, because the hole is not so visible. The shape of the numbers/letters are not clear.

ii. The main problem regarding handwritten text animation could be improved a lot so that the kids understand how they should write it, especially given with the font choice that should be suitable enough for using that effect.

iii. Regarding about the sound quality coming from the speaker, it would be better to use a camera instead of using laptop or personal computer due to the sound quality would be bad for the kids to listen. With an alternative way to get the sound right, using a phone for sound recording while using a camera for visual recording will make the quality better, including for getting a good interactive feeling.

iv. The animation wise could be better by not making it the same all the time, but should be at least different, no matter how simple it is.

v. The animation could be better by not making it the same all the time, but should be at least different, no matter how simple it is.

vi. The video is good and informative, the quality is fine, but the audio can be improved, maybe by using more prepared tools, because audio is one of the important instruments.

vii. Instead of using numbers and letters animations that uses bold and thick type of font, the application might seek other font type alternatives that gives a clearer identification and distinguish between each alphabet/number. This suggestion is based on the personal difficulty to identify zero due to the unclear shape of font and animation flow. Hence, other appropriate font type and animation could be suggested to give audience a clearer understanding.

2) Subject matter expert: There are 16 subject matter experts, including practitioners, therapists, psychologists, lecturers, and psychology students who took part in the testing phase. The subject matter expert was asked to assess the content of the animation from the medical point of view. Then the collected data will be analyzed and compiled into a bar chart.

a) Chart of Learnability for Subject Matter Expert: Fig. 9 shows the testing results from subject matter expert for the learnability aspect. The finding highlights 56.3% experts agree that the user will be able to catch the message from the host quickly as displayed in Fig. 9, this is because of the influence of good delivery and interaction from the host which means the message will be conveyed well to the users.

TABLE VI. RESULT SUMMARY FOR MULTIMEDIA EXPERT

Question type	Strongly Disagree	Disagree	Moderate	Agree	Strongly Agree	Total
Leamability	-	-	3%	55%	42%	100%
Usability	-	3%	12%	52%	33%	100%
Effectiveness	-	-	6%	67%	27%	100%
Accessibility	-	9%	15%	52%	24%	100%
Functionality	-	-	23%	50%	27%	100%
Average		2%	12%	55%	31%	100%

Fig. 9 also shows that our finding revealed that most of the respondents agreed that the application animation and content are easily understandable.

There are ten experts agreed that the instructions in the video were clear to follow as stated in Fig. 20 because the instructions from the host were very clear and delivered straight to the point without wasting time.

b) Chart of Flexibility for Subject Matter Expert: Fig. 10 shows the testing results from subject matter expert for the flexibility aspect. The finding provides evidence that eight experts agree that the content provided can promote positive behavior between children and parents as mentioned in Fig. 10. It is because it takes good cooperation and synergy between them to be able to apply the knowledge learned from the effects after watching the animated video.

Fig. 10 also presents that all experts agree that this content about learning numbers and letters is very suitable for users such as children because in general at this age the syllabus is used in the field of education in schools is general.

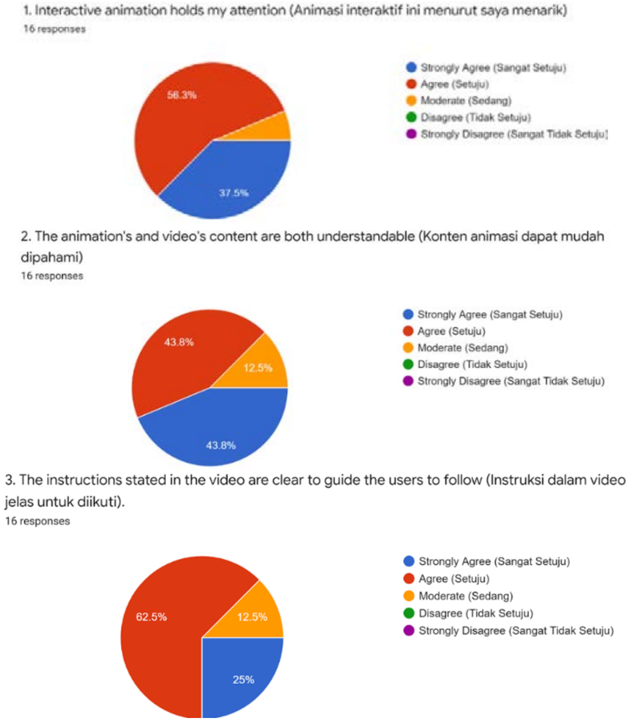


Fig. 9. Result of Learnbility by the Subject Matter Experts.

In addition, Fig. 10 stated that half of the experts agree and even the rest strongly agree that this animation will improve understanding of the content presented because children need content that is appropriate for their age such as numbers, letters, objects, colors, and so on.

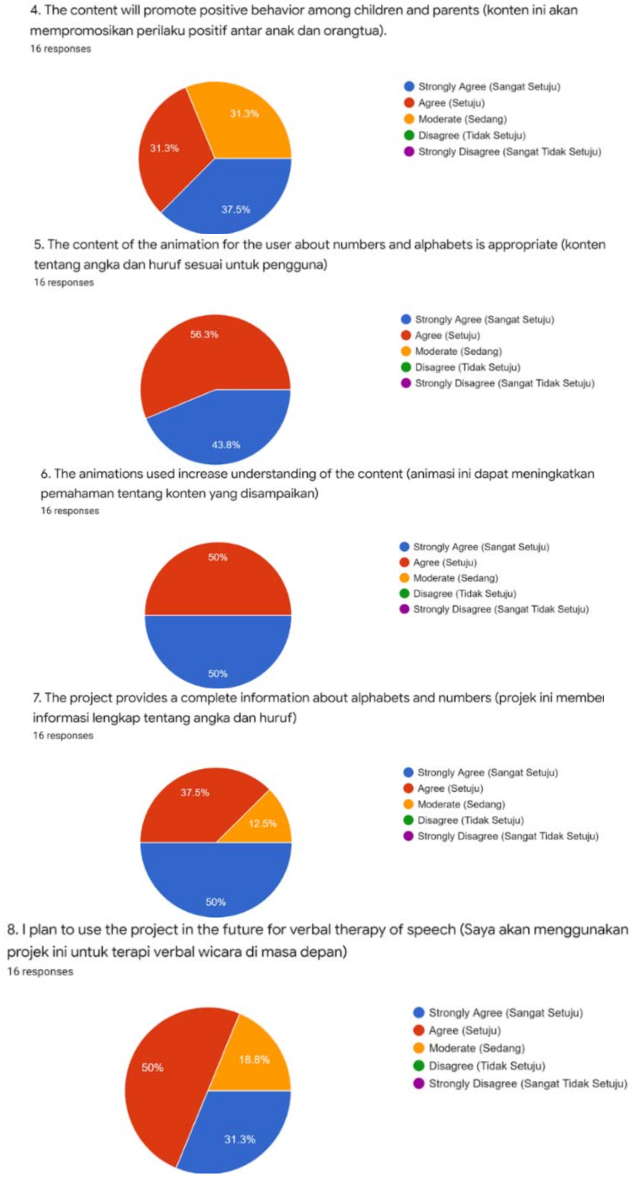


Fig. 10. Result of Flexibility by the Subject Matter Experts.

This study indicated that eight out of 16 experts strongly agree that this application provides a complete information about alphabets and numbers as mentioned in Fig. 10. This kind of syllabus is very easy to understand and remember because it will always be used in everyday life by children who watch this animated video.

Fig. 10 shows about 50% of the experts will use this application for verbal speech therapy in the future, because they believe they will be able to use it easily and without special instructions.

c) Chart of Effectiveness for Subject Matter Expert: Fig. 11 shows the testing results from subject matter expert for the effectiveness aspect. This study indicates that nine experts strongly agree that the integration of the use of multimedia elements in the animation helps users get information effectively as stated in Fig. 11 because they have used such a good, appropriate, and quite interactive elements and instruments.

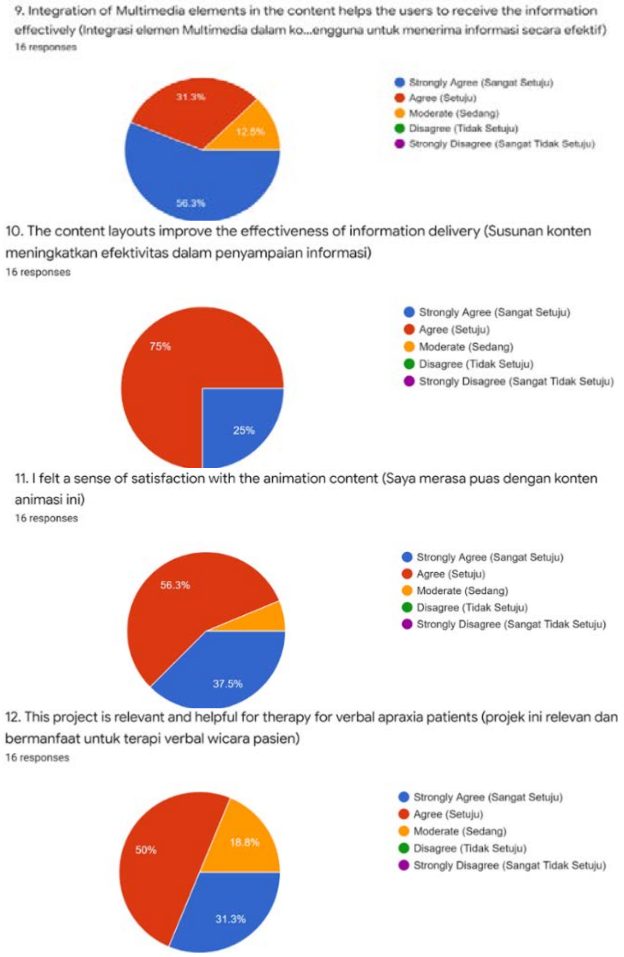


Fig. 11. Result of Effectivess by the Subject Matter Experts.

Fig. 11 presents that our finding revealed that 75% or almost all of the experts agree that the layout of this content increases the effectiveness of delivering information. This is because the arrangement is in accordance with the needs of users, namely the placement of letters, numbers, and images that are not too flashy and the selection of colors that are not monotonous makes it easy for children to remember shapes and colors different in their memory.

There are nine out of 16 experts who were satisfied, even six of them were very satisfied with this animated content as mentioned in Fig. 11. They believed it would be a good therapy for children with verbal apraxia, enabling them to learn from the animated visual interactions of the video.

Fig. 11 displays eight experts agree and five others strongly agree that this application is very relevant and believed to be useful for humans in the fields of health and education, especially for verbal speech therapy patients.

Table VII summarises test findings for subject matter experts.

d) Suggestion for Improvement from Subject Matter Expert: Based on the questionnaire provided to the subject matter expert, below are the suggestion for improvement:

i. Videos are made more colorful to help children visually receive learning.

ii. Made suitable for children to be able to guess the numbers or letters that have been learned.

iii. The material is good, only in the selection of fonts for letters, otherwise use a simple font (without an accent of beauty) so that it is easier for children to imitate.

iv. Children with special needs require less time in learning activities. A long enough duration can interfere with the child's concentration in receiving the material presented.

v. Children with disorders need more visual acuity so that animated animations can be added that can provoke cognitive thinking in children to accept learning.

vi. The background used is trying to be more dynamic, so as not to disturb the audience's focus on the back sound.

vii. This video is quite interesting, where the media used in learning is technology. As we know that children with autism disorder like new things that are not too monotonous. With interesting videos full of colors and animations, it can increase children's motivation in understanding the material presented. However, this video has a long enough duration which can interfere with the concentration of children in receiving learning. The suggestion is that this video can be divided into two sessions, where the children can first be given a break and have a moment to rest in receiving the next material.

TABLE VII. RESULT SUMMARY FOR SUBJECT MATTER EXPERT

Question type	Strongly Disagree	Disagree	Moderate	Agree	Strongly	Total
Leamability	-	-	11%	54%	35%	100%
Flexibility	-	-	12%	45%	43%	100%
Effectiveness	-	-	9%	53%	38%	100%
Average	-	-	10%	51%	39%	100%

3) Parents: There are six parents of children with verbal apraxia took part in the testing phase. They assist the testing phase to run well by explaining the questionnaire to their children with their own way.

a) Chart of Learnability for Parents: Fig. 12 shows the testing results from parents for the learnability aspect. Fig. 12 shows most parents agree and the rest strongly agree that this animation is able to attract the attention of their children because most of them acknowledge that their children are interested in interesting audio as well as cute and easy to remember animated pictures.

Half of the parents agree and the other half strongly agree that this animation is easy for their child to understand as illustrated in Fig. 12, because it contains clear instructions and keeps them focused.

There are four parents who agree and two others who strongly agree that the instructions in this video are clear for their children to follow as represented in Fig. 12 because the instructions and directions from the host are very clear with clear words and voices for the user who follows.

Fig. 12 interprets about 66.7% of parents (4 respondents) agree that learning with animation really helps their children for better understanding of numbers and letters. It is more interesting and attractive for children, as children like and are happy when there are moving pictures that resemble shapes and have eyes image to teach them how to memorize letters and numbers.

b) Chart of Effectiveness for Parents: Fig. 13 shows the testing results from parents for the effectiveness aspect. Fig. 13 depicts that four out of six chose “strongly agree” that animated interactions involve their children as a whole from the beginning of learning to completion of learning, so that children are 100% involved in the learning process through the animated visuals.

The finding highlight 100% of parents agree that this application has a good impact on their children as shows in Fig. 13. It makes children happy with the animations given and able to repeat songs and read letters and numbers, also making them easily remember the concepts learned.

Fig. 13 illustrates that our finding revealed that 66.7% of parents agree this application can promote positive behavior in children because it can make children to be more focused on something, be patient in seeing things and understand things slowly and well.

Also, Fig. 13 shows that four people agree and two others do not really agree with the statement that by using 2D animation is more effective than watching other learning videos. It is because not all children like animation, some only like the audio or the visual, not both.

c) Chart of Accessibility for Parents: Fig. 14 shows the testing results from parents for the accessibility aspect. There are five people who agree that the layout of the content increases the effectiveness in delivering information as stated in Fig. 14 because behind a neat layout, a structured pattern will appear so that the video goes according to the specified plan.

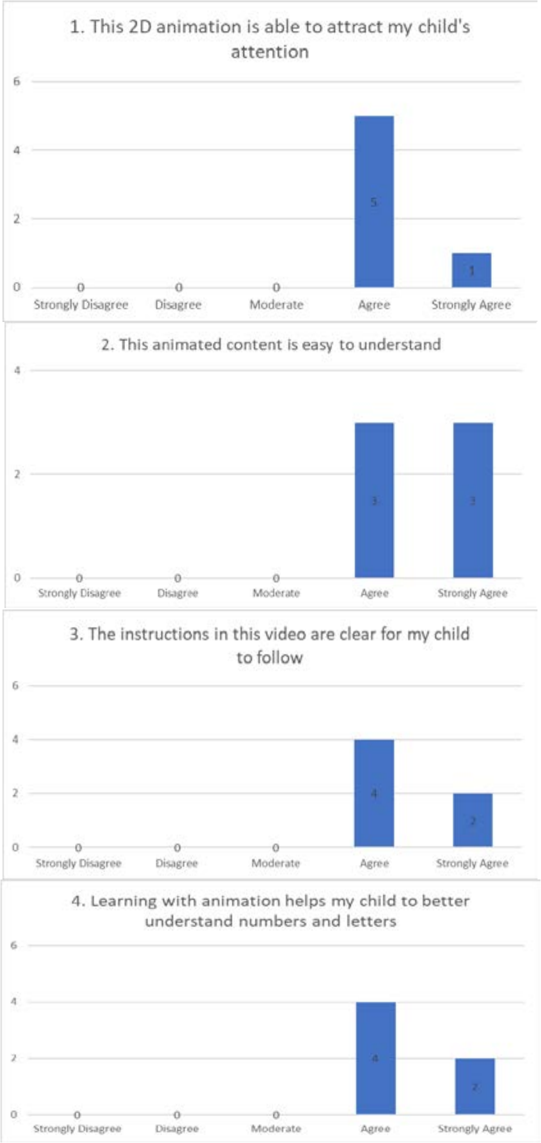


Fig. 12. Result of Learnability by the Parents.

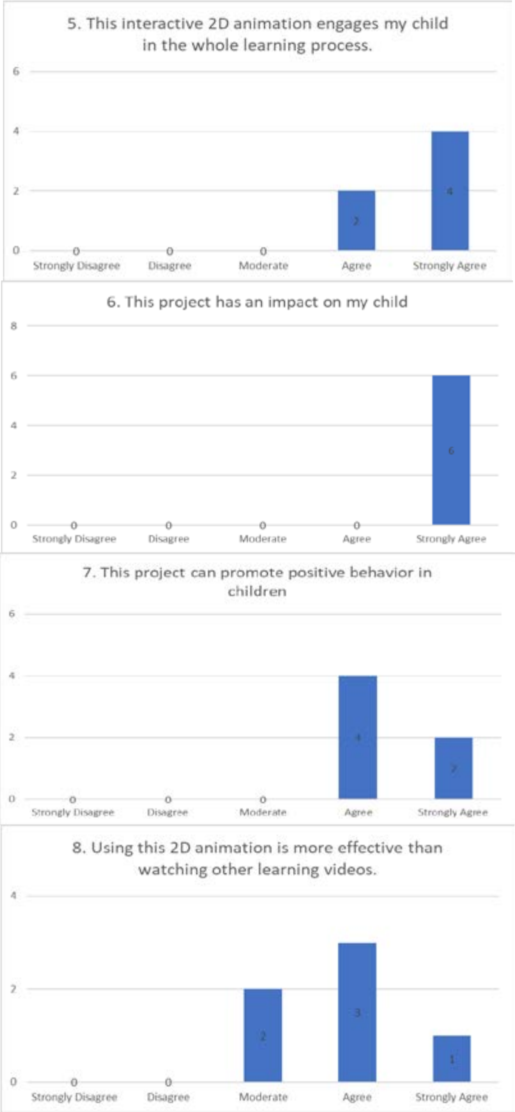


Fig. 13. Result of Effectivess by the Parents.

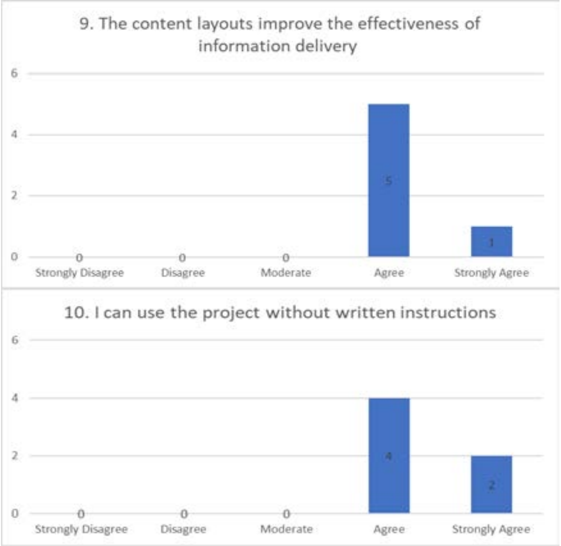


Fig. 14. Result of Accessibility by the Parents.

This study indicates that 66.7% of parents agree this application can be used easily, without written instructions as mentioned in Figure 14. This animation ran smoothly and without issue. The host and the character along with the transitions both flows well to the end. The sounds and text are both clear and smooth.

Table VIII summarises the testing summary for parents.

TABLE VIII. RESULT SUMMARY FOR PARENTS

Question type	**Strongly Disagree**	**Disagree**	**Moderate**	**Agree**	**Strongly**	**Total**
Leamability	-	-	-	67%	33%	100%
Effectiveness	-	-	8%	66%	26%	100%
Accessibility	-	-	-	75%	25%	100%
Average	-	-	8%	69%	23%	100%

d) Summary for Improvement from Parents: Based on the questionnaire distributed to the parents, below are the suggestion for improvement:

i. Videos are made more colorful to help children visually receive learning.

ii. It is suggested for this application to add the numbers up to 20 or 100, for the letters it can go up to A-Z. The rest is good and the child loves this video.

iii. Suggestion for the video, not just number and letters, but can talk about shapes, colors, animals, food and others.

iv. Quite good, and quite interesting. The child is happy with the video as the child watched it three times because the child really liked the song.

v. Do not make the video too long, suggested for two or three minutes. The rest is interesting and educational.

vi. The child may be a little annoyed with a rather long duration but enjoys.

IV. DISCUSSION

It is recommended that further improvement be undertaken such as the professional recording equipment where the equipment like voice recorders and cameras play an important role in overcoming visual and hearing problems. It is suggested to use proper equipment which can help solve the audio problem. Besides that, duration of the video also needs to be shortened which will make the audience feel less bored. Thus, dividing the topics into separate video parts and not putting them in the same video will help solve the problem. In addition, the other contents or topics of the animation should be added like types of colors, shapes, plants, human body, and other topics which is still in line with the children's learning syllabus.

The contribution of this study is obvious as the resulting outcomes can be capitalised as guidelines as verbal apraxia therapy for children with verbal apraxia of speech. The findings of this study will help demonstrate the aid for children with verbal apraxia of speech in their learning. This research work promotes children and society by providing therapy for children with verbal apraxia. It can enhance the way to conduct therapy with a new method for kids, and hopefully can be used for adults too. Furthermore, this animation includes essential knowledge that is simple to comprehend for children. This animation also provides an excellent view for children to learn about topics, such as numbers and alphabets. This application encourages children to listen to and comprehend the messages, as well as to memorize them.

V. Conclusion

The purpose of the current study was to determine the usability of the proposed application; 2D animation for Learning Numbers and Alphabets. One of the more significant findings to emerge from this study is that the proposed application received positive feedbacks from the target users. These findings suggest that in general the application is accepted and greatly contributes to the verbal apraxia children. The animation has been successfully developed and run for the target users; children with verbal apraxia. This research work also meets the objectives, which is to develop an animation video as a therapeutic solution for children with verbal apraxia, and to evaluate the effectiveness of this animation. It is very useful and could be an alternative for therapists, instructors, doctors, and parents who have children who suffer verbal apraxia, for them to learn things by a new learning method with fun and entertainment. The children are entertained by the visuals of the animation, which motivates them to learn and study new things; if this application is developed carefully with more research in accordance with the criteria recommended by medical experts, multimedia experts, and parents, this research work will have a major impact on science and education if developed carefully. The 2D animation application was developed for the society and can reach out other scopes like social, community and even science and technology.

Acknowledgment

This research is part of Final Year Project in Information and Communication Technology, presently being carried out at the Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM). We are thankful to UTeM and all respondents who took part in the final survey and testing phase, and we would like to express our thoughtful appreciation to them.

References

[1] F. Stebner, T. Kühl, T. N. Höffler, J. Wirth and P. Ayres, “The role of process information in narrations while learning with animations and static pictures,” Computers & Education, vol. 104, pp. 34-48.8, 2017.

[2] B. Sumak, A. Sorgo, “The acceptance and use of interactive whiteboards among teachers: Differences in UTAUT determinants between pre-and postadopters,” Computers in Human Behavior, vol. 64, pp. 602-620, 2016.

[3] T. Calotto and P. A. Jaques, “The effects of animated pedagogical agents in an english-as-a-foreign-language learning environment,” International Journal of Human-Computer Studies, vol. 95, pp. 15-26, 2016.

[4] Barefoot Books. “Barefoot books singalong,” accessible at https://www.youtube.com/watch?v=ynlmvWAdCug&list=PL0maGUp7cdUn5uuLnjsk1NvV2u5Yg5LAV&index=3, 2020.

[5] Terapi Austisme Online. “Terapi verbal huruf A, B, C, D, E,” accessible at https://www.youtube.com/watch?v=k3W7zQiNl6c, 2017.

[6] M. T. Hidayat, S. S. Rahim, S. Parumo, N. N. A’bas, M. A. Muhammad Sani, H. Abdul Aziz, “Designing a Two-Dimensional Animation for Verbal Apraxia for Children with Verbal Apraxia of Speech,” Ingenierie des Systemes d’Information, in press, 2022.

[7] I. Abuqaddom, H. Alazzam, A. Hudaib and F. Al-Zaghoul, “A measurable website usability model: Case Study University of Jordan,” 2019 10th International Conference on Information and Communication Systems (ICICS), pp. 83–87, 2019.

[8] R. Kaur and B. Sharma, “Comparative Study for Evaluating the Usability of Web Based Applications,” 2018 4th International Conference on Computing Sciences (ICCS), pp. 94–97, https://doi.org/10.1109/ICCS.2018.00023, 2018.

[9] J. Nielsen, “Usability 101: Introduction to Usability,” Nielsen Norman Group, https://www.nngroup.com/articles/usability-101-introduction-to-usability/, January 2012.

[10] A. Nurshuhada, R. O. M. Yusop, A. Azmi, S. A. Ismail, H. M. Sarkan and N. Kama, “Enhancing performance aspect in usability guidelines for mobile web application,” International Conference on Research and Innovation in Information Systems, ICRIIS, pp. 0–5. https://doi.org/10.1109/ICRIIS48246.2019.9073617, 2019.

[11] S. Munir, A. Rahmatullah, H. Saptono and Y. Wirani, “Usability Evaluation using NAU Method on Web Design Technique for Web Portal Development in STT Nurul Fikri,” Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019, https://doi.org/10.1109/ICIC47613.2019.8985913, 2019.

[12] K. Yamada, H. Yamana, “Effectiveness of Usability Performance Features for Web Credibility Evaluation,” Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019, pp. 6257–6259. https://doi.org/10.1109/BigData47090.2019.9006419, 2019.

[13] J. Sin, L. A. Woodham, C. Henderson, E. Williams, A. Sesé Hernández, and S. Gillard, “Usability evaluation of an eHealth intervention for family carers of individuals affected by psychosis: A mixed-method study,” Digital Health, vol. 5, https://doi.org/10.1177/2055207619871148, 2019.

[14] N. N. A’bas, S. S. Rahim, M. L. Dolhalit, W. S. N. Saifudin, N. Abdullasim, S. Parumo, R. N. R. Omar, S. Z. M. Khair, K. Kalaichelvam, and S. I. N. Izhar, “Development and Usability Testing of a Consultation System for Diabetic Retinopathy Screening,” International Journal of Advanced Computer Science and Applications, vol. 12(5), pp. 178–188. https://doi.org/10.14569/IJACSA.2021.0120522, 2021.

[15] N. N. A’bas, S. S. Rahim, M. L. Dolhalit, W. S. N. Saifudin, N. Abdullasim, S. Parumo, R. N. Raja Omar, S. Z. Md Khair, K. Kalaichelvam, and S. I. Noor Izhar, “Web Usability Testing on Diabetic Retinopathy Consultation System,” Ingénierie Des Systè Mes d ’ Information, vol. 26(3), pp. 255–264. https://doi.org/10.18280/isi.260302, 2021.

An Evaluation Method for Service-Oriented Architecture Maturity Model

Mohd Hamdi Irwan Hamzah, Ezak Fadzrin Ahmad Shaubari
Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia
86400, Parit Raja, Johor, Malaysia

***Abstract*—SOA maturity model was used to clarify and provide a common definition of SOA inside an organization. The model provides an abstract overview of SOA adoption by characterizing evolutionary levels. However, this study found that there is a lacking on how the previous models were evaluated to show that the model is conforming to the specification and can be implemented in the real-world environment. Therefore, this study aims to provide the evaluation method for the SOA maturity model through the verification and validation process. The Integrated Adoption Maturity for Service-Oriented Architecture (IAMSOA) model was chosen and the verification process is being performed through expert review where the study identifies the experts, determines the verification criteria, and collects and analyzes the feedback; while the validation was performed through case study by identifying the organization, determining the validation criteria, brainstorming, and collecting and analyzing the feedback. The verification results show that the evaluated model is comprehensive, understandable, accurate, and well-organized. Moreover, the validation results reveal that it is feasible and practical to be executed in the real environment. Conclusively, this study has successfully evaluated one of the SOA maturity models and shows the verification and validation process in detail which can be re-enacted in different projects and settings.**

Keywords—Maturity model; model evaluation; model validation; model verification; service-oriented architecture

I. Introduction

Google, Microsoft, and Facebook, among others, have grown increasingly reliant on cyberspace for everyday operations. As a result of COVID-19's rapid global spread, demand for internet services has urged in lockstep with the issue [1] [2]. The volume of service requests has increased dramatically in comparison to previous years and some of the firms have taken several significant initiatives, including the adoption of Service Oriented Architecture (SOA). A successful SOA implementation requires a well-defined SOA roadmap that details the plans, milestones, techniques, and desired outcomes. Adopting SOA is a major project that requires numerous organizational changes. Typically, businesses begin by wrapping traditional systems in web services and utilizing SOA as a means of achieving business benefits through total business transformation [3]. The most effective method for handling this shift is to divide it into phases. Adopting a maturity model is the simplest way to implement the transformation roadmap [4].

Additionally, the frequently divergent perspectives of IT, business managers, and organization leaders on SOA maturity adoption and delivery could be the result of unstated assumptions about where and why SOA should be adopted [5]. Furthermore, auxiliary work in SOA metrics is also required. Researchers are encouraged to continue developing maturity models using a combination of qualitative and quantitative metrics. A model of SOA maturity must consider both perspective and execution maturity. Progress must also be accomplished in a three-dimensional environment, with the shift from an IT-driven to an enterprise-transformation perspective — encompassing governance, performance indications, drivers, and even nomenclature, which therefore will likely eclipse execution enhancements inside a given perspective.

Nevertheless, even though numerous scholars contributed to the development of the SOA maturity model, there is still a dearth of effort in evaluating the model at the time of writing. This study has found that it is important to evaluate the model to show that it conforms to the standard specifications and is suitable to be implemented in a real-world setting. Moreover, based on the previous literature, this study has identified various SOA maturity models that have been constructed, such as SOAMM, SOASMM, Governance Maturity for SOA, and IT Risk Management Maturity Model for SOA [6, 7, 8, 9]. Yet, most of the previous models did not discuss in detail how they evaluate their models to prove they are of high quality and can produce reliable results. Based on the literature reading, the IT Risk Management Maturity Model for SOA is one of the few models that highlights the importance of evaluating the model through a case study approach. A similar finding has also been mentioned in other maturity models set up in the Global Software Development (GSD) domain, where they state that the case study approach is a powerful evaluation tool that can provide valuable real-world information [10].

However, relying solely on the case study approach to determine that the developed model conforms to its specifications and that all of the required components are present in sufficient quantity is inadequate [11]. It is critical to obtain the domain expert's approval and feedback before sanctioning the developed model [12]. Therefore, this study proposed including both the expert review (verification) and case study (validation) approach within the evaluation process. In realizing this approach, it was determined that one of the SOA maturity models should be evaluated to determine whether the constructed model meets the specifications, captures the requirements needed, and can be implemented in a real-world environment. Accordingly, the Integrated Adoption Maturity for Service-Oriented Architecture (IAMSOA) model was chosen and evaluated through the verification and validation approaches. This study is structured as follows: Section 1 covers the introduction and is followed by Section II (the background of the study). Section III discusses the methodology applied in this study and Section IV presents the results. Section V discusses the verification and validation results and Section VI summarizes the study.

II. Background of the Study

SOA maturity models provide an abstract overview of SOA adoption by characterizing evolutionary levels [13] [14]. They can be thought of as a collection of critical process areas that work together to optimize a well-defined business and IT architecture. They also can be used to regulate and measure the progress of SOA adoption [15]. In addition, SOA maturity is one of the most pressing challenges of SOA adoption issues in the SOA lifecycle [16]. However, theoretically sound, methodologically rigorous and empirically evaluated SOA maturity models are extremely rare [17]. As a result, this study determined the importance of reviewing and discussing many of the most widely used SOA maturity models presented by the industry and academia. The aim of the study is to determine how the current researchers evaluate their proposed SOA maturity model.

The SOA Maturity Model (SOAMM) was published in 2005 based on the feedback from around 2000 architects and developers [6]. SOAMM maturity levels were evaluated through a single evaluation dimension that incorporates various perspectives, including prime business benefits, scope, critical technology success factors, critical people success factors, organizational success factors, and selected relevant standards. Based on these perspectives, they presented a guide for establishing SOA vision and a benchmark for measuring progress by including the goals, characterization of the scope, business benefits, important industry standard, key practices, and critical success factor. Despite this, SOAMM never explained how they evaluate their model in depth.

In 2012, Welke et al. [14] proposed an SOA maturity model based on the capabilities of maturity model integration (CMMI). Welke's model first interpreted their maturity model as a capability orientation model, and then they specified that as SOA becomes more mature, the SOA ability should be fully realized in order to contribute to business operations and organization's service orientation as a whole. Furthermore, the researcher proposed an SOA maturity cube that introduces the idea of a multidimensional view for SOA maturity. The first dimension is for the organization to identify their current levels of SOA maturity according to six defined SOA criteria, which are the infrastructure efficiency, reuse, composition and integration, business process and analytics, enterprise flexibility and agility and enterprise transformation; the second dimension is to determine what to do in order to reach the next maturity, which are the benefits and metrics, business involvement, methodology, service sourcing and governances. However, their study also does not include a comprehensive discussion on the evaluation method.

Inspired by the introduction of the SOA maturity model in 2005 [6] and the maturity cube concept by Welke et al. [14], Hamzah et al. [18] first presented that the SOA should have prioritized on both information technology (IT) and business benefits. Then, continuing their work, they mentioned that the SOA maturity level should be based on the Adoption of Innovation theory to cater for SOA adoption issues and that the GQM approach should have been included within the evaluation matrix to provide a structure and systematic evaluation [19] [20]. Additionally, they did an exploratory study in 2019 to bolster and enhance their findings, which subsequently lead to the introduction of the IAMSOA model. Nonetheless, their work is incomplete, as they emphasize the importance of evaluating their model to conform that it can accomplish the desired objective and being employed in a real-world setting.

Coincidentally, within the same year of 2019, another researcher has performed a systematic literature review on the SOA maturity models to identify research opportunities and areas where the SOA maturity model can be improved [21]. A total of 20 unique SOA Maturity Models were investigated and reviewed in detail. Their findings reveal that although all SOA Maturity Models propose an assessment framework, only a few SOA Maturity Models are guided in prioritizing the improvement process. Furthermore, in line with this study's interest, they also mentioned that empirical research on the in-depth analysis and evaluation of SOA Maturity Models is sparse. They also further acknowledged that there is minimal effort in evaluating the SOA maturity model and it is critical for future research to work on this issue.

Recently in 2021, Azevedo et al. [9] constructed an IT Risk Management Maturity Model for SOA. This work presents a risk management maturity model, formed by the union of good information technology risk management practices and existing maturity models, to be applied in an SOA. The model aims to support the assessment process of identifying the level of risk management maturity within the SOA domain. To evaluate the proposed model, the scenario of a health organization was used, and the results indicated that the level of IT risk management maturity based on SOA was measured, providing a holistic view of risk management on the dimensions of people, processes, and technology. They stated unequivocally that it is important to validate the SOA maturity model through a real-world scenario in which their results show the risks maturity level and the importance of managing risks properly.

As a result of earlier literatures, this study concluded that it is important to perform the evaluation on the SOA maturity model, particularly through the implementation in a real-world environment. The proposed evaluation method through the verification and validation process is expected to provide an organized guideline for other researcher to evaluate their developed SOA maturity model. This study will build on the work of [20] by conducting an evaluation of the IAMSOA model via the verification and validation process. The approach utilized in this study to evaluate the IAMSOA model is described in the next section.

III. THE EVALUATION METHOD FOR SERVICE-ORIENTED ARCHITECTURE MATURITY MODEL

Based on the literature reading, this study has found that the evaluation for SOA maturity model should be performed through two main phases which are the verification and validation phases. Fig. 1 shows the proposed evaluation method for SOA maturity model and the components require to perform the evaluation process. They are discussed further in the next subsection.

A. Verification Stage

The verification should be performed to check whether the SOA maturity model conforms to its specification [22] and ensures that all required components are present in the right quantity [23]. This study found that the verification stage was intended to verify i) the maturity level, ii) evaluation dimension, and iii) evaluation matrix. To accomplish these aims, the expert review should be used because it has been accepted as a significant way to detect and remove defects [24]. Basically, there are three activities involved in verifying the SOA maturity model which are i) identifying the expert, ii) determining the verification criteria, and iii) analyzing the feedback. These activities are discussed in the following subsections.

1) Identifying the expert: The experts should be chosen among the academicians (knowledge experts) by following the characteristics of experts as suggested by Hallowell and Gambatese [25]. The characteristics include i) currently attached to the field of the study under examination, ii) hold an advanced degree (PhD.), iii) faculty members at an accredited university, iv) authorship, and v) have at least five years of experience. Additionally, as the SOA maturity model is intended to be used by the SOA practitioners, therefore, they should be included as the domain experts to perform the verification as well as to give their insights from the real-life environment point of view. The characteristics of the domain experts are that they should have at least three years of experience in SOA implementation.

2) Determine the verification criteria: The study has identified that the major components of the SOA maturity model are the maturity level, evaluation dimension, and the evaluation matrix. These components should be verified for their comprehensiveness, understandability, accurateness, and organization. These criteria were based on the previous studies where it was appropriate and has been successfully used to verify their model or framework [26] [27]. The questions that were asked to verify and measure these criteria were adopted from Salah [28].

3) Collecting and analyze the feedback: The knowledge expert's feedbacks should be collected and analyzed for further improvements.

B. Validation Stage

Models validation is a fundamental process to confirm that the models are of sufficiently high quality [29]. In this context, the validation process was performed to prove that the SOA maturity model have high quality and can be re-enacted in other projects or settings. This study performed the validation process through the case study approach. The details activities for the validation process are discussed in the next subsection.

1) Identifying the organization: The organization should be selected based on their dealing with SOA, the available projects related to the SOA, and their willingness to apply the model. The study also identified that it is appropriate to select the organization according to their expertise in dealing with SOA where the organization needs to be competent in SOA-based applications. This enables the testing of the feasibility and practicality of the model in different settings.

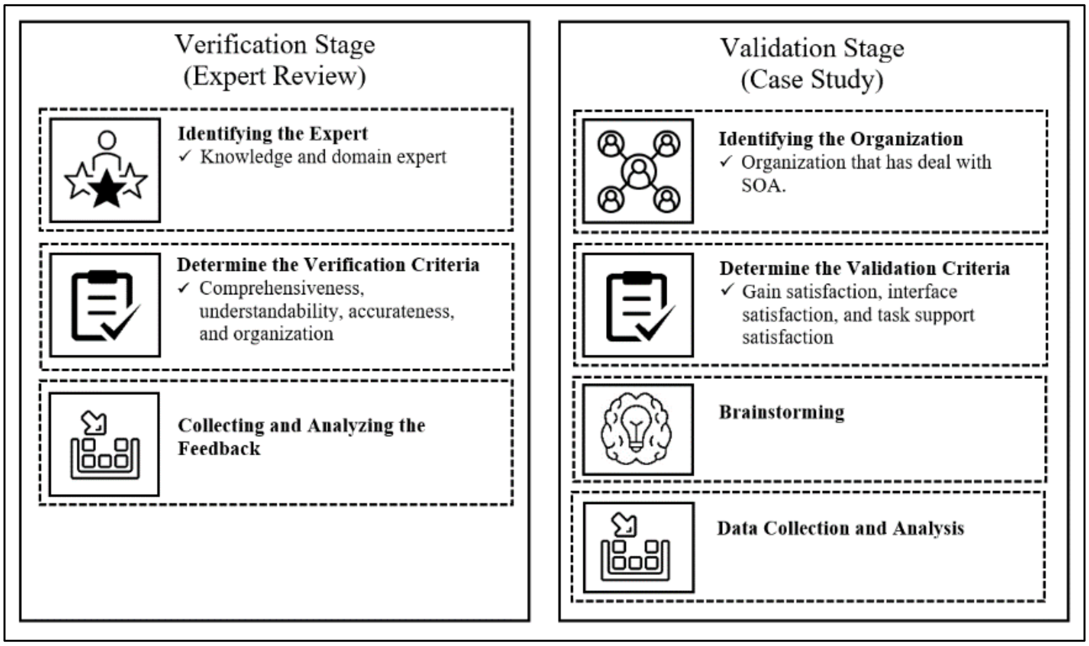

Verification Stage
(Expert Review)
Identifying the Expert
✓ Knowledge and domain expert
Determine the Verification Criteria
✓ Comprehensiveness, understandability, accurateness, and organization
Collecting and Analyzing the Feedback
Validation Stage
(Case Study)
Identifying the Organization
✓ Organization that has deal with SOA.
Determine the Validation Criteria
✓ Gain satisfaction, interface satisfaction, and task support satisfaction
Brainstorming
Data Collection and Analysis

Fig. 1. The Proposed Evaluation Method for SOA Maturity Model.

2) Determine the validation criteria: The validation criteria for the SOA maturity model were determined by adapting them from the study of [25] which can reveal the success of the proposed model. Kitchenham and Pickard [30] stated that the evaluation criteria should include three main criteria which are, gain satisfaction, interface satisfaction, and task support satisfaction. These common criteria had then been used by several researchers such as [20] and [19] in evaluating their model or framework which were carried out in the field of software engineering.

3) Brainstorming: After the identification of the validation criteria, the validation process should be performed. The brainstorming session should be conducted where it is to introduce the SOA maturity model to the organization. The purpose of implementing the SOA maturity model should be explained to the organization who participate in this validation process.

4) Data collection and analysis: The data collection and analysis can be performed by giving out and collecting the data based on two evaluation forms which are: i) the proposed instrument in order to evaluate the maturity of SOA adoption for the organization and ii) the evaluation form to validate the proposed model. Based on the feedback from the organization, the data should be analyzed and the evaluated SOA maturity model should be improved.

IV. The Implementation of the Proposed Evaluation Method for Service-Oriented Architecture Maturity Model

This section is going to discuss the result of implementing the proposed evaluation method tos the IAMSOA model.

A. *Result for Verification*

This section illustrates the experts' answers and suggestions for the IAMSOA verification. There are three major components of the IAMSOA model which need to be verified by the experts which are the evaluation dimension, maturity level, and evaluation matrix. These components were verified based on their comprehensiveness, understandability, accurateness, and organization. These criteria were adapted from previous studies [24] [26].

The experts provided their feedbacks by filling in the checklist form. The experts were asked to rank the level of these criteria achievement. The Six Likert scales were used to describe the level of achievement of the items. The results were calculated by getting the mean score for each criterion and selecting the appropriate interval that represents the actual mean. Table I shows the mean interval presentation and the achievement level adapted from ISO 15504 while Table II reveals the verification results for the evaluation dimension components.

Results in Table II show that three out of four criteria gained "fully achieved" for the evaluation dimension which are comprehensiveness, understandability, and well-organized. The accuracy is the only criteria that gained 'largely achieved'. Most of the experts stated that the evaluation dimension is well defined and acceptable.

TABLE I. Representation of Achievement Levels (Adapted from ISO/IEC 15504)

Mean Interval Presentation	Achievement level
From 0 to 0.8 (0%-15%)	Not achieved
From 0.9 to 2.9 (>15%-50%)	Partially achieved
From 3 to 5 (>50%-85%)	Largely achieved
From 5.1 to 6 (>85%-100%)	Fully achieved

TABLE II. Verification Results for the IAMSOA Evaluation Dimension

Item	Mean	Overall Mean	Achievement Level
Comprehensiveness			
The required criteria for evaluating the SOA IT and business benefits are included.	5.3	5.3	Fully Achieved
The required sub-criteria for evaluating the SOA IT and business benefits are included.	5.3		
Accuracy			
The IT and business benefits criteria and sub-criteria are correctly assigned to maturity levels.	4.9	5	Largely Achieved
There is no overlap detected for the descriptions of IT and business benefits criteria and sub-criteria.	5.1		
The sub-criteria for IT and business benefits are correctly assigned to IT and business benefits criteria.	5		
Understandability			
The IT and business benefits criteria and sub-criteria are understandable.	5.2	5.2	Fully Achieved
The IT and business benefits criteria and sub-criteria descriptions are understandable.	5.2		
Well-Organized			
The IT and business benefits criteria are well organized.	5.3	5.25	Fully Achieved
The IT and business benefits sub-criteria are well organized.	5.2		

Expert F mentioned that at the 'Optimized' level, the organization should be looking at the agility and flexibility for optimization and transformation to be ahead of the competition. Furthermore, another expert (Expert C) stated that the proposed model needs to consider a few KPAs such as Configuration Management. In short, all of the experts agreed with the evaluation dimension of the IAMSOA by stating that the proposed dimension is acceptable and can be applied to measure the IT and business benefits. As for the IAMSOA maturity level verification, the results are listed in Table III.

TABLE III. VERIFICATION RESULTS FOR THE IAMSOA MATURITY LEVEL

Item	Mean	Overall Mean	Achievement Level
Comprehensiveness			
The number of maturity levels are adequate and appropriate.	5.2	5.08	Fully Achieved
The maturity levels description is sufficient.	5.3		
The maturity levels are sufficient to represent all maturation stages of the domain.	5		
The Key Process Area (KPA) for each maturity level covers all aspects for evaluation of the domain.	4.8		
Accuracy			
There is no overlap detected between descriptions of maturity levels.	5	5.13	Fully Achieved
There is no overlap detected between each Key Process Area (KPA) of maturity levels.	5		
The Key Process Areas (KPAs) are correctly assigned to their respective maturity level.	5.4		
Understandability			
The maturity levels are understandable.	5.3	5.2	Fully Achieved
The maturity levels description are understandable.	5.2		
The Key Process Areas (KPAs) for each maturity level are understandable	5.1		
Well-Organized			
The maturity levels are well organized.	5.2	5.15	Fully Achieved
The Key Process Area (KPA) for each maturity level are well organized.	5.1		

As shown in Table III, the maturity level gained "fully achieved" for all the criteria. Based on the expert review, majority of the experts satisfied with the proposed maturity level. One of the experts mentioned that the maturity level for the IAMSOA model is derived from a well-defined standard in software engineering practices such as CMMI and it is well understood. Another expert also stated that the maturity level is well organized and understandable. Conclusively, all the experts agreed that the IAMSOA maturity level is well-defined, organized, and appropriate to be used as a benchmark for measuring the SOA maturity and adoption. Table IV presents the results of the IAMSOA evaluation matrix verification.

TABLE IV. VERIFICATION RESULTS FOR THE IAMSOA EVALUATION MATRIX

Item	Mean	Overall Mean	Achievement Level
Comprehensiveness			
The model is sufficient to determine the SOA adoption maturity	5	4.92	Largely Achieved
The model is sufficient for conducting SOA adoption process improvement	5		
The model is sufficient to track the SOA adoption issues	4.7		
The model is sufficient and practical to be used in industry	4.9		
The model is useful to be used in industry	5		
Accuracy			
The overall evaluation matrix are constructed correctly	5.1	4.98	Largely Achieved
The evaluation goals are constructed correctly	5		
The evaluation questions are constructed correctly	4.9		
The evaluation metrics are constructed correctly	5.1		
The evaluation process are constructed correctly	4.8		
Understandability			
The evaluation matrix is understandable	5.3	5.2	Fully Achieved
The evaluation process is understandable	5.3		
The implementation guideline is understandable	5		
The scoring scheme is understandable	5.2		
Well-Organized			
The structure of evaluation matrix are well organized	5.4	5.3	Fully Achieved
The evaluation process are well organized	5.2		

Based on Table IV, two out of four criteria for the IAMSOA evaluation dimension gained "fully achieved". The criteria that obtained "fully achieved" are understandability and well-organized, while the comprehensiveness and accuracy criteria "largely achieved". Expert E commented that detail descriptions for the achievement scale (not achieved, partially achieved, largely achieved, and fully achieved) are required and the coverage for the evaluation needs to be included. Moreover, Expert A mentioned that the evaluation matrix needs to include specific steps such as an evaluation flow for the evaluation matrix to be clearer and understandable. Another expert (Expert C) also stated that there are a few steps in the assessment process that need to be rearranged where some of the steps should appear at the beginning of the assessment process. Nevertheless, despite some of their comments, all the experts agreed that the evaluation matrix for IAMSOA is well structured and can be applied in the real-world environment.

Overall, all experts gave general reviews of the proposed IAMSOA model. They concluded that IAMSOA is a good model and can benefit the industry. Majority of the experts agreed that IAMSOA is a flexible model that can be extended to cater other related SOA domains. Finally, once the evaluation dimension, maturity level, and evaluation matrix have been verified, the validation process were performed and is presented in the next subsection.

B. *Result for Validation*

The aim of the validation process is to evaluate the effectiveness of the IAMSOA model. In this study, the validation was performed through case study. This section presents the results of the case study conducted in one of the software companies in Malaysia. The aim is to validate the IAMSOA model and show its applicability and added benefits. Fig. 2 shows the assessment process flow for SOA maturity model.

The case study was performed by assessing the Product Inspection System by Company A. Detail discussions for the case study are presented in the following subsection.

1) Organization profile: Company A is an in-house solution provider. Its main client is an electric utility company that represents one of the large semi-government sectors in Malaysia. Apart from providing a centralized, one-stop center for technical solutions and innovation, Company A has developed some solutions under a variety of applied research projects. Recently, Company A needed to ensure that only qualified equipment is employed by their parent company. In order to provide this service, a Product Inspection (PI) Management System was developed by the company. This signifies that the PI has an enormous job coordination. To handle such big volumes, a flexible architecture is necessary to ensure that the PI Section is able to manage various PI related works. Hence, the SOA has been adopted in developing the PI Management System. In addition, a mechanism to measure the quality of the SOA-based application is required to ensure the delivery of high-quality services to the IT and business people in the company. Having the SOA-based application, it is obvious that the IAMSOA model can help in determining the level of adoption and maturity. After contacting and discussing with the company's ICT Research Unit, the proposed IAMSOA model was used to assess the SOA adoption maturity for the PI Management System. The subsequent sections describe the details of the IAMSOA model related processes in evaluating and assessing the SOA adoption maturity in the company.

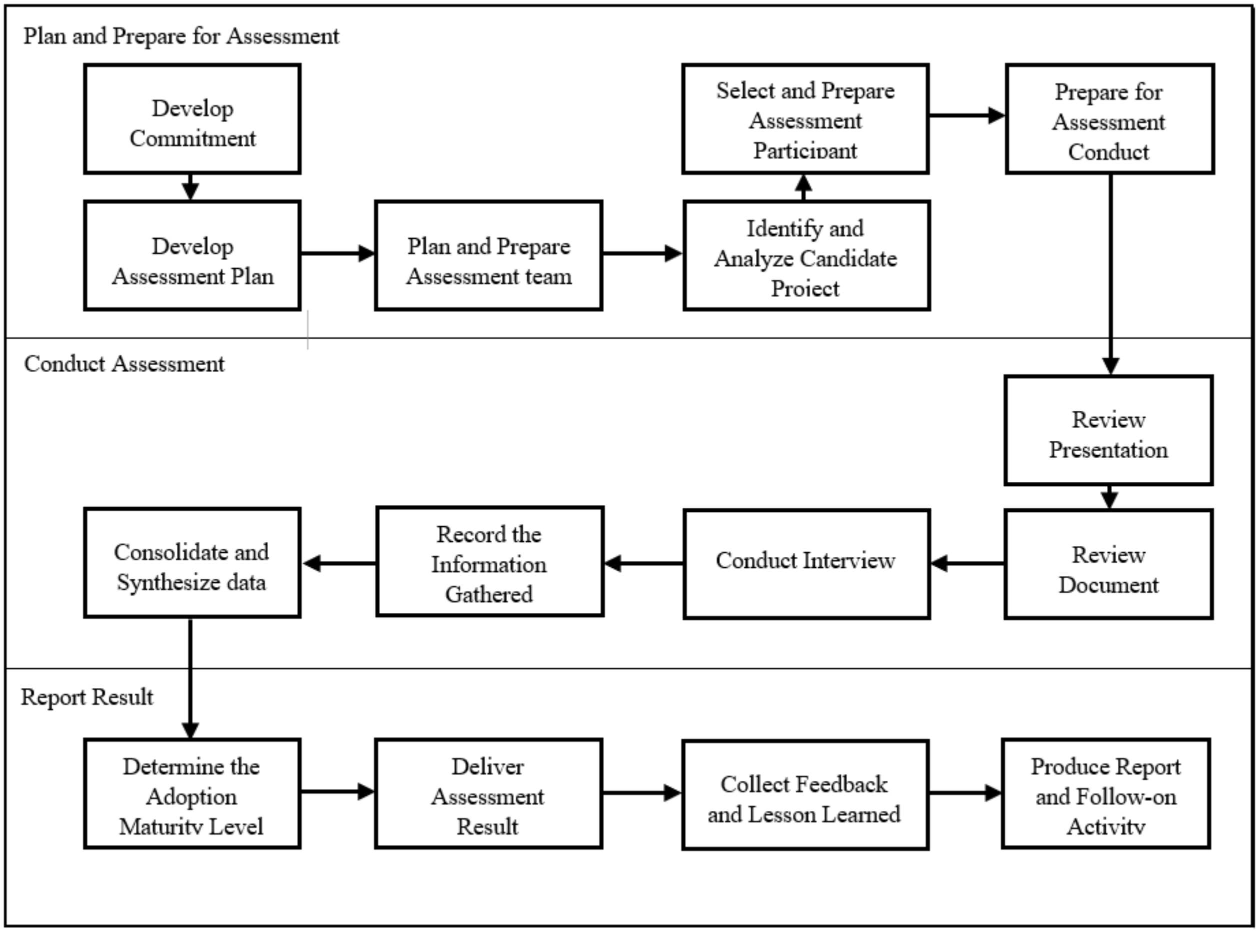

Plan and Prepare for Assessment
Develop Commitment
Develop Assessment Plan
Plan and Prepare Assessment team
Identify and Analyze Candidate Project
Select and Prepare Assessment Participant
Prepare for Assessment Conduct
Conduct Assessment
Review Presentation
Review Document
Conduct Interview
Record the Information Gathered
Consolidate and Synthesize data
Report Result
Determine the Adoption Maturity Level
Deliver Assessment Result
Collect Feedback and Lesson Learned
Produce Report and Follow-on Activity

Fig. 2. SOA Maturity Model Assessment Process Flow.

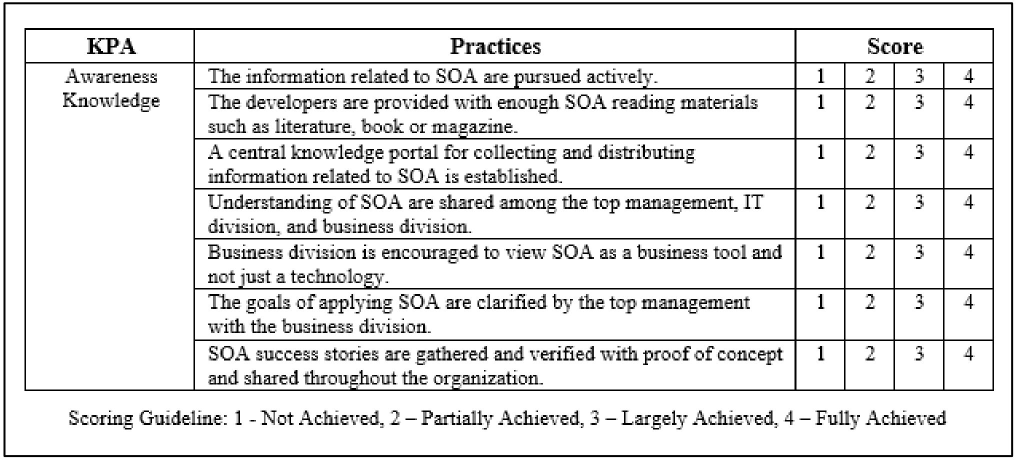
2) Plan and prepare for assessment process: There are six activities involved in the planning and preparing for the assessment process; i) developing commitment, ii) developing assessment plan, iii) planning and preparing assessment team, iv) identifying and analyzing project candidate, v) selecting and preparing assessment participant, and vi) preparing assessment conduct. During this phase, the assessment team in Company A started to plan and prepare for applying the IAMSOA model. The phase started by defining the objectives, constraint, and scope for appropriate assessment design as well as establishing an organization leader commitment. Then, the team prepared a document for guiding and defining the execution of the assessment. The lead assessor, conducted a briefing session to the team to familiarize them with the assessment plan, IAMSOA model, and assessment process. After the briefing, the team continued with a meeting among its members to discuss whether the assessment can be conducted or not. After they agreed to proceed with the assessment, each staff was appointed a specific role.

3) Conduct assessment process: After finishing the first activity, the assessment team continued to conduct the assessment. The assessment processes include i) reviewing presentation, ii) reviewing document, iii) conducting interview, iv) recording the gathered information, and v) consolidating and synthesizing data.

This phase began when the assessor started to collect the data by participating during the staff/developer's presentation regarding SOA project. The assessor then assessed the documents produced during the development of the project. The assessor also interviewed the top management, project manager, and developers to obtain information about the project/system and clarify on any information that could not be acquired through document review. Then, the assessor produced a document that records all the gathered information. After obtaining all the required data, the assessor went on to calculate the score for individual and overall performance qualities of the SOA project by using the IAMSOA assessment form as shown in Fig. 3.

4) Report result process: The final phase is about reporting the results that involved four activities; i) determining the adoption maturity level, ii) delivering assessment results, iii) collecting feedbacks and lesson learned, and iv) producing report and supporting follow-on activity. At the beginning of this phase, the SOA adoption maturity level for Company A was determined based on the score extracted from the IAMSOA assessment form. The results of the maturity level for the PI Management System is presented in Tables V, VI, VII, and VIII. Table V shows the score for the PI Management System at Maturity Level 1 is 95% which indicates that the PI Management System has successfully achieved maturity level 1.

Table VI shows the score for the PI Management System at Maturity Level 2 of 91.5%. This indicates that the PI Management System has successfully achieved maturity level 2.



KPA	Practices	Score			
Awareness Knowledge	The information related to SOA are pursued actively.	1	2	3	4
	The developers are provided with enough SOA reading materials such as literature, book or magazine.	1	2	3	4
	A central knowledge portal for collecting and distributing information related to SOA is established.	1	2	3	4
	Understanding of SOA are shared among the top management, IT division, and business division.	1	2	3	4
	Business division is encouraged to view SOA as a business tool and not just a technology.	1	2	3	4
	The goals of applying SOA are clarified by the top management with the business division.	1	2	3	4
	SOA success stories are gathered and verified with proof of concept and shared throughout the organization.	1	2	3	4

Scoring Guideline: 1 - Not Achieved, 2 – Partially Achieved, 3 – Largely Achieved, 4 – Fully Achieved

Fig. 3. Example of the IAMSOA Model Assessment Form.

TABLE V. MATURITY LEVEL 1 FOR THE PI MANAGEMENT SYSTEM

Key Process Area	Key Practices	Individual Quality	Overall Quality	Achievement
SOA Knowledge Gathering (SOAKG)	Awareness Knowledge (AK)	85%	95%	>85% Achieved Maturity Level 1
	How-to Knowledge (HK)	90%		
	Principle Knowledge (PK)	100%		
New Functionality (NF)	Perform New Service (PNS)	100%		
	Develop Pilot Project (DPP)	100%		

TABLE VI. MATURITY LEVEL 2 FOR THE PI MANAGEMENT SYSTEM

Key Process Area	**Key Practices**	**Individual Quality**	**Overall Quality**	**chievement**
SOA Adoption (SOAA)	SOA Adoption Decision (SOAAD)	100%	91.5%	>85% Achieved Maturity Level 2
	SOA Infrastructure Management (SOAIM)	85%		
	SOA Best Practices Management (SOABPM)	95%		
	SOA Project Planning (SOAPP)	87%		
Service Integration (SI)	Service Modularity (SM)	95%		
Service Scalability (SS)	Service Migration (SMi)	87%		
Cost Reduction (CR)	Time Management (TM)	87%		
	Cost Management (CM)	96%		

Table VII shows the score for the PI Management System at Maturity Level 3 of 93.5%. This shows that the PI Management System has successfully achieved maturity level 3.

TABLE VII. MATURITY LEVEL 3 FOR THE PI MANAGEMENT SYSTEM

Key Process Area	**Key Practices**	**Individual Quality**	**Overall Quality**	**Achievement**
SOA Implementation (SOAI)	Technical Assistance Resolution (TAR)	96%	93.5%	>85% Achieved Maturity Level 3
	Service Analysis (SA)	87%		
	Service Design (SDES)	89%		
	Service Development (SDEV)	93%		
	Service Monitoring (SMo)	100%		
Service Reusability (SR)	Service Publicity (SP)	100%		
	Service Conformance (SC)	100%		
	Service Comprehensibility (SCo)	100%		
	Service Understandability (SU)	100%		
Service Integration (SI)	Service Availability (SAv)	91%		
Service Flexibility (SF)	Service Reliability (SRe)	91%		
Service Agility (SA)	Service Modifiability (SMod)	100%		
Service Scalability (SS)	Service Replication (SRe)	100%		
IT/Business Alignment (ITBA)	Orchestration Management (OM)	91%		
	Resources Alignment (RA)	85%		

Table VIII shows the score for the PI Management System at Maturity Level 4 of 77.4%. This indicates that the PI Management System has not achieved maturity level 4.

TABLE VIII. MATURITY LEVEL 4 FOR THE PI MANAGEMENT SYSTEM

Key Process Area	**Key Practices**	**Individual Quality**	**Overall Quality**	**Achievement**
SOA Performance Evaluation (SOAPE)	Service Level Agreement (SLA)	95%	77.4%	<85% Not Achieved Maturity Level 4
	System Testing (ST)	100%		
Service Reusability (SR)	Service Discoverability (SD)	83%		
	Service Commonality (SCom)	92%		
	Service Composability (SComp)	75%		
	Service Portability (SP)	62.5%		
	Service Adaptability (SAd)	62.5%		
Service Flexibility (SF)	Service Interoperability (SInt)	42%		
	Service Changeability (SCh)	55%		
Service Agility (SA)	Service Evolvability (SEv)	75%		
Business Quality (BQ)	QoS Assurance (QoSA)	95%		
	Security Management (SM)	92%		

Results from the evaluation show that the PI Management System has achieved "Maturity Level 3". Maturity level 5 Key Process Area was marked with 'pending' once the Company A failed to achieve Maturity Level 4. Based on Table VIII, Company A can instantly identify at which Key Process Areas and Key Practices that they have been lacking off. Therefore, it is important for Company A to be able to measure the relative maturity within each Key Process Area to identify areas that are lacking.

As presented in Table VIII, at maturity level 4, the PI Management System partially achieved the Service Flexibility, and largely achieved the Service Reusability and Service Agility. Thus, Company A needs to give more attention and work on these Key Process Areas to achieve the next level which is Maturity Level 4. Once the lagging Key Process Areas have been identified, it is possible to come up with solutions and eventually improve the success of the overall SOA initiative.

Based on the results presented in Tables V, VI, VII, and VIII, Company A can determine which goal that the PI Management System has already achieved according to the IAMSOA model. The goal achievement is important to identify the areas that are already fulfilled and those that need improvement. After determining the SOA adoption maturity, the assessment results were immediately presented to the company's ICT Research Unit to get an agreement on the outcomes. In addition, based on the results, recommendation on future improvements were proposed. Company A has to improve three main service areas, namely Flexibility, Reusability, and Agility in order to achieve Maturity Level 4 and to progress to the next level Maturity Level 5.

V. Discussion

The results indicate that Company A has achieved maturity level 3 and implies that Company A is competent in applying the SOA based-application. Moreover, an interview session was conducted with the assessment team leader for Company A to validate the IAMSOA model. The team leader answered the validation form that was constructed based on a set of evaluation factors. These factors are gain, interface, and task support satisfactions. Each factor includes various related items or statements. These statements were answered by the assessment team members by deciding whether to AGREE or DISAGREE. This type of answers format signifies a practical measurement [31] that can directly capture the respondent's intention effectively. Table IX displays the validation results form.

From Table IX, the first criterion or factor to be evaluated during the validation process is "gain satisfaction" that measures the benefit of the IAMSOA model to the real-life environment. The measurement items for this factor include decision support satisfaction, comparison with the previous model, clarity, and task appropriateness. The results from the interview point out that the assessment team from company A stressed that the model achieved decision support satisfaction by helping the organization to decide on well-defined processes. The assessment team also agree that the model is very clear and understandable where each process presents the required input, outputs, methods and activities.

TABLE IX. Validation Results

Item	Company A
Gain satisfaction	
Decision support satisfaction: The IAMSOA model helps the management to take a well-defined decision based on the processes.	Agree
Comparison with the previous SOA model: The IAMSOA is better than the old model that you used in terms of structure and achieved results.	Agree
Clarity (clear and illuminate the process): The IAMSOA process is clear to the development team, where each phase clearly presents the required inputs, outputs, methods or practices, and activities.	Agree
Task Appropriateness: The phases and activities presented in the IAMSOA model are appropriate for adopting and implementing SOA in your company; and the flow of the process is presented in a systematic and effective way.	Agree
Interface satisfaction	
Internally consistent: The IAMSOA model is internally consistent.	Agree
Organization (well organized): The components of the IAMSOA model are well organized and structured which makes the process easy to perform.	Agree
Appropriate for audience: The IAMSOA model is appropriate for the audience. Those audiences are referred to the development and the monitoring team in the software firms.	Agree
Presentation: The results presented by performing the IAMSOA process are produced in a readable and useful format.	Agree
Task support satisfaction	
Ability to produce expected results: The IAMSOA model is able to produce expected results.	Agree
Completeness (adequate or sufficient): The IAMSOA model is adequate and sufficient for adopting and implementing SOA in your organization.	Agree
Ease of implementation: The process of the IAMSOA model is easy to implement.	Agree
Perceived usefulness	
Using IAMSOA model enables you to accomplish your tasks more quickly.	Agree
Using IAMSOA model improve the performance of your work.	Agree
Using IAMSOA model makes performing your tasks easier.	Agree
IAMSOA model is useful to your work.	Agree
Using IAMSOA model increases your productivity.	Agree
Perceived ease of use	
Learning the IAMSOA model is easy for you.	Agree
Do you find it easy to use IAMSOA model to do what want to do?	Agree
The IAMSOA model is flexible to interact with.	Agree
Your interactions with the IAMSOA model are clear and understandable.	Agree
It is easy for you to become skillful in using the IAMSOA model.	Agree
The IAMSOA model is easy to use.	Agree

The “interface satisfaction” represents the second criterion or factor that measures the IAMSOA model in terms of interface presentation, format, and processing efficiency. Pertaining to this factor the assessment teams emphasized that the model is internally consistent whereby each component complements one another. The team also agreed that the IAMSOA components are well organized and structured by sorting all the processes, activities, and roles in a clear and understandable manner. The model is also declared to be appropriate for the audience as the team members comprised of those with variety of skills. The team also satisfied with the readable and useful results format produced based on the IAMSOA model.

To ensure that the IAMSOA model can achieve its intended purpose and satisfies the assessor, the “task support satisfaction” factor is used as the third measurement factor. The team also agreed that the IAMSOA model can produce the expected result because it provides a well-defined sequence of activities and a wide variety of evaluation criteria such as the SOA adoption, maturity, IT, and business benefits. In addition, the model was found to be adequate and sufficient in determining the level of SOA adoption maturity that focuses on the IT and business benefits. Consequently, based on the responses relating to perceived usefulness and ease of use, it can be concluded that the IAMSOA model is useful, easy to use, effective, and feasible to be used in Malaysian organizations.

VI. Conclusion

This study has presented the evaluation method for the SOA maturity model through the verification and validation stages. The significance of the proposed evaluation method within the domain of the study is to provide a guideline for other researchers who want to ensure that their SOA maturity model conforms to the specification and can be implemented in the real-world environment. The IAMSOA model was chosen as a case study to implement the proposed evaluation method for the SOA maturity model. In the verification stage, the IAMSOA model has been verified by five knowledge and five domain experts. The experts were asked to verify the IAMSOA evaluation dimension, maturity level, and evaluation matrix. Next, the validation stage was performed using the case study approach, and the findings reveal that the IAMSOA model is feasible to be implemented in the real world based on several real-life applicability evaluation criteria such as gain satisfaction, interface satisfaction, task support satisfaction, perceived usefulness, and perceived ease of use. Conclusively, this study has successfully evaluated one of the SOA maturity models through the verification and validation process. Moreover, this study has contributed to explaining how to perform the expert review and case study to evaluate the SOA maturity model. The detailed process, methods, results, and discussion have been presented, and this can guide other researchers to perform the evaluation of their models in other settings.

Nevertheless, this study also identified that there is a limitation to the proposed evaluation method, which requires future work towards enhancing this study. The limitation is that the proposed method involves collaborative assessment where organizations form a team to assess the maturity of their SOA-based applications. Based on the assessment results, Company A achieved maturity level 3 in their first assessment. The achievement of maturity level 3 for the first assessment is considered high compared to the CMMI assessment (third-party assessment). The high achievement might relate to the collaborative assessment whereby the assessment was performed by its own staff, which may cause biased assessment. Another reason may relate to the practices for the assessment being too general. Thus, in the future, a third-party assessor can be brought in, and the practices should be changed to make the results more reliable and high-quality.

Acknowledgment

The authors would like to thank the Ministry of Higher Education Malaysia for supporting this research under Fundamental Research Grant Scheme Vot No. FRGS/1/2020/ICT03/UTHM/03/1 and partially sponsored by Universiti Tun Hussein Onn Malaysia.

References

[1] V. Thuy-Anh, M. Mieszko, and T. An, “The impact of COVID-19 economic crisis on the speed of adjustment toward target leverage ratio: An international analysis,” Finance Research Letters, vol. 45, 2021, https://doi.org/10.1016/j.frl.2021.102157.

[2] A. Savanevicien, G. Radvila, and V. Šilingien, “Structural changes of organizational maturity during the COVID-19 Pandemic: The Case of Lithuania,” Sustainability, vol. 13, 2021, https://doi.org/10.3390/su132413978.

[3] T. Catarci, D. Firmani, F. Leotta, F. Mandreoli, M. Mecella, and F. Sapio, “A conceptual architecture and model for smart manufacturing relying on service-based digital twins,” in IEEE International Conference on Web Services (ICWS), 2019, pp. 229-236, doi: 10.1109/ICWS.2019.00047.

[4] W. Aniruddha Anil, J. Rohit, R. Ajay Pal Singh, and J. Rakes, “Development of maturity model for assessing the implementation of Industry 4.0: learning from theory and practice,” Production Planning and Control, vol. 32, no. 8, pp. 603-622, 2021, https://doi.org/10.1080/09537287.2020.1744763.

[5] N. Niknejad, A. R. C. Hussin, and I. S. Amiri, “Literature review of service-oriented architecture (SOA) adoption researches and the related significant factors. In: The impact of service oriented architecture adoption on organizations,” in SpringerBriefs in Electrical and Computer Engineering. Springer, Cham, 2019, https://doi.org/10.1007/978-3-030-12100-6_2.

[6] Sonic Software Corporation, AmberPoint Inc, BearingPoint Inc, and Systinet, A new Service-Oriented Architecture (SOA) maturity model, 2005. [Online]. Available: http://www.omg.org/soa/Uploaded Docs/SOA/SOA_Maturity.pdf, 2005.

[7] M. Kassou, and L. Kjiri, "SOASMM: A novel service oriented architecture Security Maturity Model," International Conference on Multimedia Computing and Systems, 2012, pp. 912-918, doi: 10.1109/ICMCS.2012.6320279.

[8] M. Dehghani, and S. Emadi, "Developing a New Model for Governance Maturity of Service Oriented Architecture," Soft Computing Journal, vol. 4, no. 2, pp. 54-67, 2021.

[9] R. Azevedo, and P. Caetano, “An Approach of Risk Maturity Models for SOA,” in International Workshop on ADVANCEs in ICT Infrastructures and Services (ADVANCE 2021), 2021, pp. 3-12.

[10] N. Rashid, S. U. Khan, H. U. Khan, and M. Ilyas, "Green-Agile Maturity Model: An Evaluation Framework for Global Software Development Vendors," in IEEE Access, vol. 9, pp. 71868-71886, 2021, doi: 10.1109/ACCESS.2021.3079194.

[11] M. Alaa, I. S. M. A. Albakri, C. K. S. Singh, H. Hammed, A. A. Zaidan, B. B. Zaidan, O. S. Albahri, M. A. Alsalem, M. M. Salih, E. M. Almahdi, M. J. Baqer, N. S. Jalood, S. Nidhal, A. H. Shareef, and A. N. Jasim, "Assessment and Ranking Framework for the English Skills of

Pre-Service Teachers Based on Fuzzy Delphi and TOPSIS Methods," in IEEE Access, 2019, vol. 7, pp. 126201-126223, 2019, doi: 10.1109/ACCESS.2019.2936898.

[12] C. Adrian, R. Abdullah, R. Atan, and Y. Y. Jusoh, "Expert Review on Big Data Analytics Implementation Model in Data-driven Decision-Making," 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), 2018, pp. 1-5, doi: 10.1109/INFRKM.2018.8464770.

[13] Y. Baghdadi, “SOA maturity models: Guidance to realize SOA,” International Journal of Computer and Communication Engineering, vol. 3. no. 5, pp. 372-378, 2014. http://doi.org/10.7763/IJCCE.2014.V3.352.

[14] R. Welke, R. Hirschheim, and A. Schwarz, “Service-oriented architecture maturity,” in IEEE Computer Society, 2011, vol. 44, pp. 61-67, http://doi.org/10.1109/MC.2011.56.

[15] M. Niemann, C., Janiesch, N. Repp, and R. Steinmetz, “Challenges of governance approaches for service-oriented architectures,” in: Proceedings of the Third IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST, 2009), 2009, pp. 600-605.

[16] E. Hustad, and D. H. Olsen, “Creating a sustainable digital infrastructure: the role of service-oriented architecture,” Procedia Computer Science, vol. 181, no. 1, pp. 597–604, 2021.

[17] L. Lasrado, and R. K. Vatrapu, “Maturity models development in IS research: A literature review,” in Proceedings of the 38th Information Systems Research Seminar in Scandinavia (IRIS 38), 2015, (August 9-12). http://doi.org/10.13140/RG.2.1.3046.3209.

[18] M. H. I. Hamzah, F. Baharom, and H. Mohd, “Constructing service-oriented architecture adoption maturity matrix using Kano model,” in International Conference on Applied Science and Technology (ICAST), 2017. https://doi.org/10.1063/1.5005382.

[19] M. H. I. Hamzah, F. Baharom, and H. Mohd, “A construction of service-oriented architecture adoption maturity levels using adoption of innovation concept and CMMI,” Journal of Telecommunication, Electronic and Computer Engineering, vol. 10, pp. 23-27, 2018.

[20] M. H. I. Hamzah, F. Baharom, and H. Mohd, “Evaluation of service-oriented architecture adoption maturity model for sustainable development,” Journal of Telecommunication, Electronic and Computer Engineering, vol. 10, pp. 2-4, 2018.

[21] S. Pulparambil, and Y. Baghdadi, Service oriented architecture maturity models: A systematic literature review, in Computer Standards & Interfaces, 2019, vol. 61, pp. 65-76, ISSN 0920-5489, https://doi.org/10.1016/j.csi.2018.05.001.

[22] M. Chemuturi, Mastering software quality assurance. Florida: J.Ross Publishing, 2011.

[23] I. Sommerville, Software engineering. 10th Ed. Pearson India, 2018.

[24] M. Komuro, and N. Komoda, “An explanation model for quality improvement effect of peer reviews,” in International Conference on Computational Intelligence for Modelling Control & Automation, 2008, pp 1159-1164, doi: 10.1109/CIMCA.2008.187.

[25] M. R. Hallowell, and J. A. Gambatese, “Qualitative research: application of the delphi method to CEM research,” Journal of construction engineering and management, 2010, vol. 136, no. 1, pp. 99-107. doi: 10.1061/_ASCE_CO.1943-7862.0000137.

[26] F. H. Al-Tarawneh, “A framework for cots software evaluation and selection for cots mismatches handling and non-functional requirements,” Ph.D. dissertation, School of Computing, UUM, Sintok, 2014.

[27] M. Shafinah Farvin Packeer, F. Baharom, A. Deraman, J. Yahya, and H. Mohd, “An exploratory study on secure software practices among software practitioners in Malaysia,” Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 2016, vol. 8, no. 8, pp. 39–45.

[28] D. Salah, R. Paige, and P. Cairns, “An evaluation template for expert review of maturity models,” vol. 8892, pp 318–321, 2014.

[29] M. B. Carvalho, F. Bellotti, J. Hu, J. B. Hauge, R. Berta, A. Gloria, and M. Rauterberg, “Towards a service-oriented architecture framework for educational serious games,” in IEEE International Conference on Advanced Learning Technologies (ICALT), 2015, vol. 4, pp. 147-151, http://doi.org/10.1109/ICALT.2015.145.

[30] B. Kitchenham, “Evaluating software engineering methods and tool,” in ACM SIGSOFT software engineering Notes, 1998, vol. 23, no. 5, pp. 21-24.

[31] B. M. McSkimming, S. Mackay, and A. Decker, "Investigating the usage of Likert-style items within Computer Science Education Research Instruments," in 2021 IEEE Frontiers in Education Conference (FIE), 2021, pp. 1-8, doi: 10.1109/FIE49875.2021.9637198.

Fake Face Generator: Generating Fake Human Faces using GAN

Md. Mahiuddin[1], Md. Khaliluzzaman[2], Md. Shahnur Azad Chowdhury[3], Muhammed Nazmul Arefin[4]
Dept. of Computer Science and Engineering, International Islamic University Chittagong (IIUC) [1, 2, 4]
Chattogram- 4318, Bangladesh[1, 2, 4]
Dept. of Computer Science and Engineering, Chittagong University of Engineering and Technology[1]
Chattogram- 4349, Bangladesh[1]
Dept. of Business Administration, International Islamic University Chittagong (IIUC), Chattogram- 4318, Bangladesh[3]

***Abstract*—As machine learning is growing rapidly, creating art and images by machine is one of the most trending topics in current time. It has enormous applications in our current day to day life. Various researchers have researched this topic and they try to implement various ideas and most of them are based on CNN or other tools. The aim of our work is to generate comparatively better real-life fake human faces with low computational power and without any external image classifier, rather than removing all the noise and maximizing the stabilization which was the main challenge of the previous related works. For that, in this paper, we tried to implement our generative adversarial network with two fully connected sequential models, one as a generator and another as a discriminator. Our generator is trainable which gets random data and tries to create fake human faces. On the other hand, our discriminator gets data from the CelebA dataset and tries to detect that the images generated by the generator are fake or real, and gives feedback to the generator. Based on the feedback the generator improves its model and tries to generate more realistic images.**

Keywords—Generative adversarial network; fake human faces; generator; discriminator; CelebA dataset

I. Introduction

Fake images can be used by police to track down people who are involved in adultery. As we know nowadays child and underage pornography is alarming. To catch such people we can use computer generated fake faces, that will be more ethical than using real images [11]. As we can see nowadays creating fake or edited images is very easy. To identify these kinds of images, we can build a model. For this model to be trained we can use the computer generated images to identify the real and fake images. We use the fake computer generated images in visual art and the advertising industries. These images have a huge role in computer games. In computer and mobile games along with play station games, these images can be used. As augmented reality and virtual reality are growing rapidly, we can use these computer generated fake images.

II. Literature Review

There are many fake image generation models created or updated by the researchers which generate pretty good images. Zhang et al. [1] in 2020 discussed an algorithm that can improve the quality of generated images as the quality of the fake images generated by the conventional GAN is limited. Wang et al. [2] in 2016 tried to simplify the overall process of generative models which gave them more realistic high resolution images as well as highly stable and robust learning procedures. Ghatas et al. [3] in 2020 proposed a method of building a complex modular pipeline using previously trained models to generate the kin image. They tried to build such a model where it changes the way of approaching GAN problems.

Hamdi et al. [4] in 2019 proposed a new GAN that uses the K-nearest neighbor for selective feature matching in a high level space and they named their work as K-GAN. Tolosana et al. [5] in 2020 tried to recognize fake news, fake images, fake media, and face manipulation using their method. They provided a thorough review of digital manipulation techniques which were applied to facial content due to the huge number of possible detrimental applications.

Karras et al. [6] in 2019 proposed a method that re-designs the generator architecture so that we can understand the various aspects of the image synthesis process. Though GAN has seen rapid improvement in recent years, we still have a very poor understanding of the properties of the latent spaces. They tried to understand the inner workings of the latent spaces. Their work starts learning from constant input and tries to adjust the “Style” of the image on each layer. Their method was tested on Flickr-Faces-HQ and FFHQ datasets.

Zhao et al. [7] in 2019 proposed an image translation network by exploiting attributes with the generative adversarial network. It can remarkably contribute to the seven authenticity of the generated face by supplementing the sketch image with the additional facial attribute feature. The generator and discriminator both use skip-connection to reduce the number of layers without affecting network performance. In the underlying feature extraction phase this network is different from most attribute-embedded networks. They divided their networks into two parts as sub-branch A and sub-branch B, which takes a sketch image and attribute vector to extract low level profile information and high level semantic features.

In recent years, Generative Adversarial Networks have achieved extraordinary results for various applications in many fields especially in image generation because of their ability to create sharp and realistic images. In this paper, Shirin Nasr Esfahani et al. [8] discussed five areas of image synthesis based on different techniques of GAN. They are: Text to Image Synthesis, Image to Image translation, Face Manipulation, 3D Image Synthesis and Deep Master Prints. In this paper they have actually tried to focus on the applications of the above techniques and discussed their merits and demerits and future applications. An Introduction to Image Synthesis with Generative Adversarial Nets is proposed in [10]. This model is working well for the simple dataset, for the complex dataset the model does not work well.

Though there are many fake image generation models, what we have tried to do in our work is that we have tried to generate the images using low computational power and without any external image classifier. We have used two neural networks in our model. One is Generator and the other is Discriminator. Generator creates fake human faces and Discriminator detects the faces generated by the generator if this is fake or real. If the face generated by the generator is detected as fake then the discriminator gives feedback to the generator. And based on the feedback of the discriminator, the generator improves itself and tries to generate better quality faces. This process goes on until the generator creates better quality fake human faces. From detecting fake news, fake images to data misclassification, fake human face generation helps us to be used in all of these areas. And generative adversarial network model is one of the best one to create better real life fake human faces.

Rest of the paper is organized as follows. The proposed method is explained in Section III. In Section IV, the experimental results are presented. The paper is concluded in Section V.

III. Proposed Method

A. Methodology

In our model we have tried to implement generative adversarial networks in a very classic way with new and latest tools and technologies. We have tried to implement a generative adversarial network to generate fake human faces from random noise with low computational power and without any external image classifier, rather than removing all the noise and maximizing the stabilization which was the main challenge of the previous related works. We know that for implementing generative adversarial network we have to set a generator whose work will be to generate fake faces from the noise and update its own model after receiving the feedback from the discriminator and we also have to implement a proper discriminator to identify the real and fake image, so that it give the right feedback, means it says true to real image and false to fake image and based on its feedback the generator will change its model in such a way that eventually it will create almost real like images.

That’s how we are trying to implement our generative adversarial network to generate fake human faces. Our generator and discriminator will use the keras’s classic sequential modeling.

We proposed to use two sequential models, one as a generator and another one as a discriminator to implement our generative adversarial model. For implementing the generator, as we can see from the workflow diagram, we have used random noise as input data for the generator. We have also used the Batch Normalization technique and the Reshape module from the keras to resize the data and normalize to train the generator properly. In our generator we have used a total of five layers, one is an input layer with 128 nodes, starting with three more dense layers 512, 1024 and 2048 nodes respectively with each layer and lastly the output layer. We have used the Leaky ReLU activation function for all the layers except the output layer; in the output layer we have used the tanh activation function for the output layer. We have also used the loss function as the binary cross-entropy for our generator model. Our generator will receive feedback data from the discriminator and based on the received feedback our generator will be updated.

On the other hand, our discriminator is also implemented with the sequential model. However, the main difference between our generators is that after each epoch it does not learn, meaning it will not update or train after each epoch. Our discriminator’s main purpose is to identify the real and fake image and give the feedback to the generator, so that we will use the sequential model. In our discriminator’s input data we have used both the generator produced data and the real image from the CelebA dataset and it takes both images and tries to identify the real and the fake image. Here, the model uses a total of four layers. As an input layer we have taken the generated data and the real image data and there are also two hidden dense layers along with an output layer which produce binary values. Other than the output layer we have used the Leaky ReLU activation function in our model and as our output will be classification type, so in the output layer we have used the sigmoid function. For the loss function in our model we have used the binary cross-entropy function.

B. Overall Workflow

In Fig. 1 the diagram we have shown the overall workflow of our proposed model, and later in this paper we have also explained how our generator, discriminator and the fully connected sequential model work [12] [13].

C. Generator

We have proposed two sequential models, one as a generator and another one as a discriminator to implement the generative adversarial model. We have used random noise as input data in the generator model. The input data is actually a layer which has 100 nodes. Then we have used four hidden layers where we have used Leaky Relu as our activation function. The hidden layers have 128, 256, 512, 1024 nodes respectively. We have used the Tanh activation function in our last layer. We have also used Batch Normalization to normalize the data. At last, we have reshaped the nodes into (48, 48, 1) size. The Adam optimizer has been used as our optimizer. During the compilation of the model, binary cross-entropy has been used as our loss function. The generator will produce some fake human faces and it will improve itself by the feedback from the discriminator. The flow diagram of the generator is presented at Fig. 2.

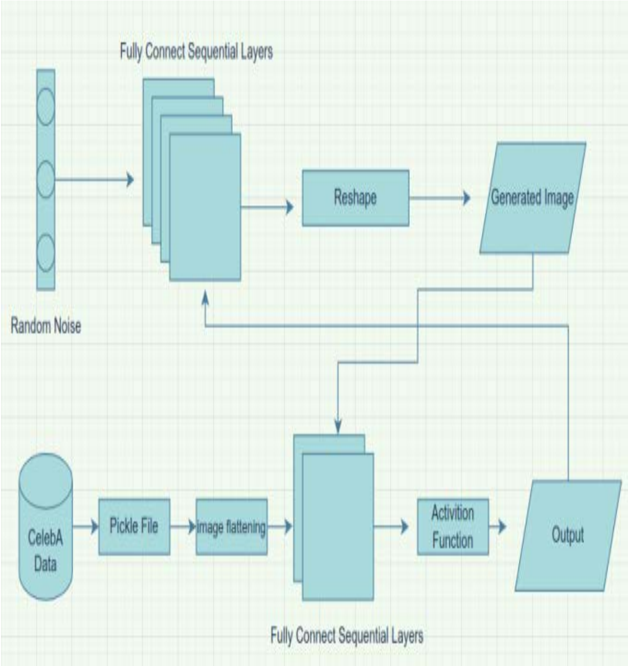


Fig. 1. Overall Workflow Diagram.

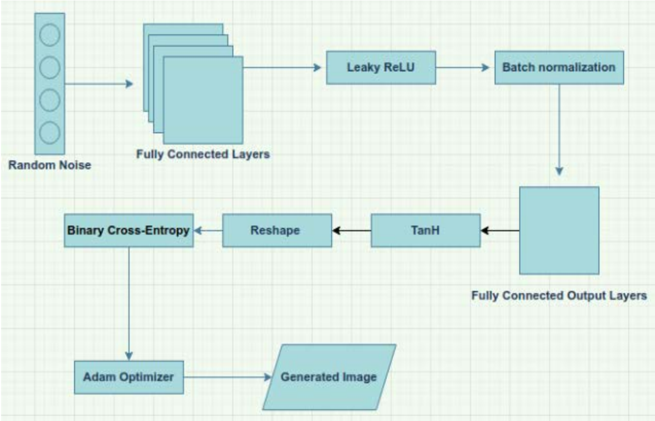


Fig. 2. Generator.

D. Discriminator

Discriminator is our second sequential model which we have used in our generative adversarial model. At first we used the flatten operation to make the data one dimensional. After that there are two dense layers. The first hidden layer has 48*48 nodes and the second hidden layer has half of the first hidden layer, which means 1152 nodes. Leaky ReLU has been used as the activation function in those two dense layers having the learning rate as 0.2. In the output layer the sigmoid activation function has been used as the activation function. We have used binary cross-entropy as the loss function and Adam as the optimizer during the compilation period. The discriminator will receive images from both the actual data from the celebA dataset and fake data from the generator and it will give feedback to the generator if the images generated by the generator are fake or real. The generator will receive feedback from the discriminator and improve itself. The flow diagram of the discriminator is presented in Fig. 3.

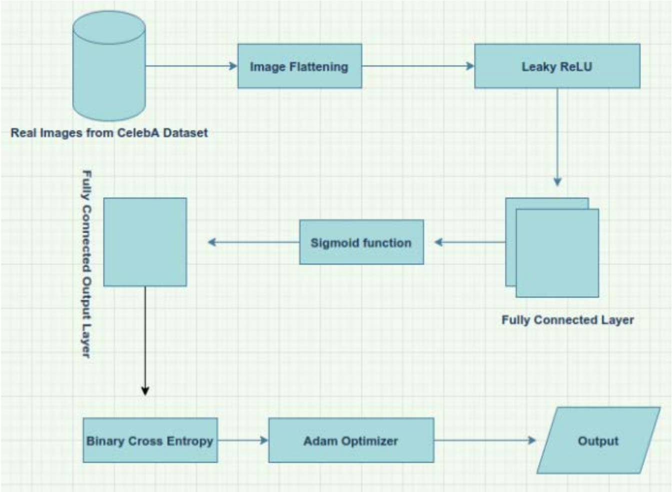


Fig. 3. Discriminator.

In our proposed model we have used two fully connected sequential models, one as a generator and another one as a discriminator. The generator receives a random noise and tries to train it as a human face and then our discriminator receives data from the CelebA dataset and tries to match the faces with the generator's fake faces, and give feedback to the generator, and based on the feedback our generator learns and tries to create comparatively better fake human faces after each training.

IV. Experimental Results

In this section, we have presented a detailed overview of the experiments of our presented model based on the CelebA dataset. We have shown how our model worked during different scenarios of the dataset. We have changed the size of the data multiple times to see how the model works each time. And we have got different results each time. We have described how the results differ and improve over the time.

A. Dataset and Experimental Settings

We have used the CelebA dataset for our model. In our CelebA dataset there are more than 200k data. At first, we ran our model on this whole dataset but we used the size of images as 48*48. We ran 100000 epochs on the model to get a good result. Normally in generative adversarial models we have to run a lot of epochs to get a good result. Generative adversarial models generally generate new images according to the task given. So, it requires a lot of time to generate the new images. That's why we ran the model 100k epochs [14].

B. Experimental Tools and Environment

We have used different tools and environments for preprocessing the data and running the model so that we can have good results from our model.

C. Programming Language

Tensorflow and Keras, which are the two libraries of the Python programming language, have been used in our model. Nowadays Python is the most popular programming language for implementing Machine Learning and Deep Learning models as it has many collections of packages which are very helpful to implement the different Machine Learning and Deep Learning models. Python is also the most preferred language because it is very easy to learn and implement. Python has many free and open source libraries for Machine Learning and Deep Learning models. Tensorflow is one of them. Keras is an open source software library which supports a Python interface for artificial neural networks. Keras plays the role as an interface for the Tensorflow library.

D. Result

The four images shown in the Fig. 4 were generated by the same model after 52000, 65000, 81000 and 97000 epochs respectively. In the first image two faces could not be generated properly. Rest of the faces of the first image were generated but still there were a lot of noises which actually prevented us from understanding the faces properly.

Compared to the first one, the second image is slightly a little bit better but there are also one or two faces which could not be generated properly and contained a lot of noise. Almost all of the images were generated properly in the third image containing some noise, though it was comparatively better than the first two images. In the fourth image, this image performed quite well. In the fourth image all of the faces were generated and had a small noise. After receiving the images from the CelebA database we first flatten the images then we run the sequential model with the Leaky ReLU. Then in the output layer we use the sigmoid function to return the true or false type result value. Then in the discriminator we use the binary cross entropy as the loss function along with the Adam optimizer. This output from the discriminator then goes to the generator as a feedback and based on the feedback of the discriminator, the generator tries to create new images. So, after the above discussion we can say that there are still noises in the generated faces at the end of the model too, though the results of the latter part of the model performed comparatively better.

The four images shown in Fig. 5 were generated after 52000, 63500, 81500 and 99500 epochs respectively. After the 52000 epochs, two faces could not be generated properly, which we can see from the first image above. Some faces are not perfect though some faces are generated properly with little noise. In the second image which was generated after 63500 epochs, all of the faces were almost properly generated with some noise. After 81500 epochs and 99500 epochs, the faces which were generated, that we can see from the above third and fourth images, are much better. The interesting fact is that if we observe the faces properly we can see that almost all of the faces are different from one another.

After the above discussion what we can observe is that the faces became better as the time went on. And at the last phase of the model the faces became more and more clear and understandable. Another interesting fact is that as the time went on all of the faces were generated by which we can understand that the model was improving over time. And we can also see the faces were different from one another which is also a very positive output of this model.

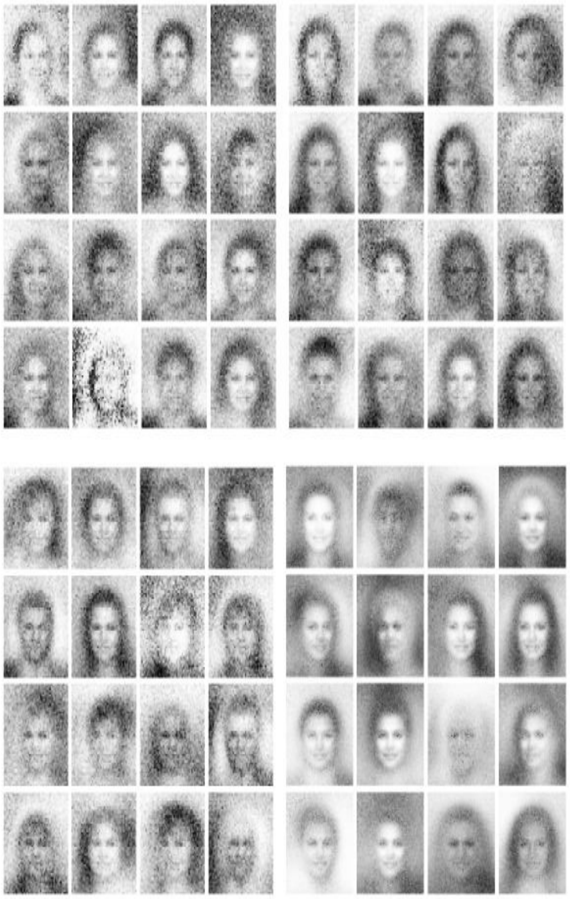


Fig. 4. Generated Data of 48*48 Size (CelebA Dataset).

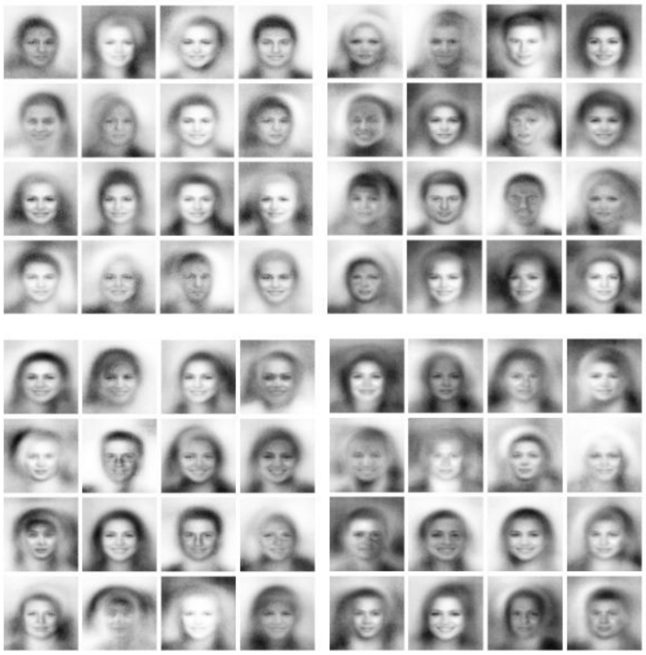


Fig. 5. Generated Data of 100*100 Size (CelebA Dataset).

E. Loss Function

Here, in the Fig. 6, we can see the loss function of the discriminator. We ran our models at about 100000 epochs. If we see the graph we can see most of the loss value between 0 to 1, which is pretty good value for the discriminator. And as the discriminator is an un-trainable model we see the loss value maintains the same ranges between the whole 100000 epochs. Now we can see some spike in the loss value of our graph. If we try to separate the high spike by trying again and again to visualize the diagram without the high spike of the loss value in our (48 x 48) size images, we can then easily see our loss value is somewhere between 0 and 1.4. And if we calculate the high spike, we can see there are 771 high spikes between 100000 epochs which is around 0.77% high spike and other 99.23% are the average, means 0 to 1.4 ranges. This is pretty good value for our model.

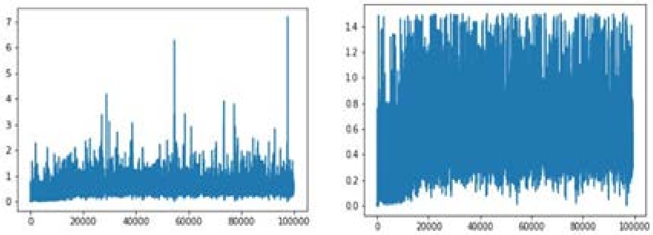


Fig. 6. Loss Values.

F. Comparative Analysis

Zeliha Dogan [9] et al. proposed a model titled “Baby Face Generation with Generative Neural Networks”, where they tried to generate fake baby faces using comparatively low computational power. We have chosen this paper for comparison because both this paper and our work follow almost the same objectives. Both this paper and our model mainly tried to generate fake faces using comparatively low computation power. Below in this section we have tried to compare our work with this paper, and show where our proposed method stands out and why. We have also tried to show our drawbacks and possible ways to overcome this drawback.

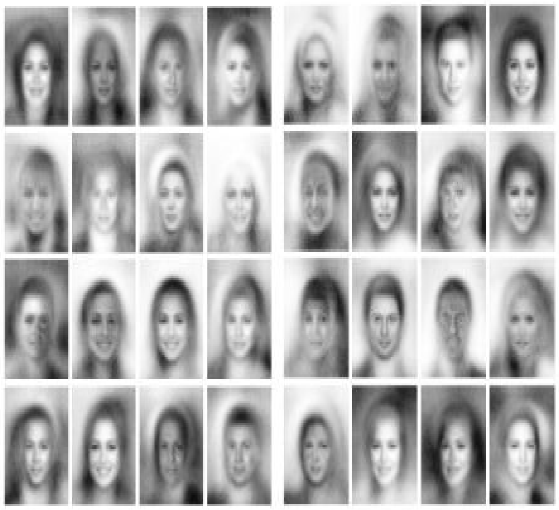


Fig. 7. Sample Output Images of the Proposed Model.

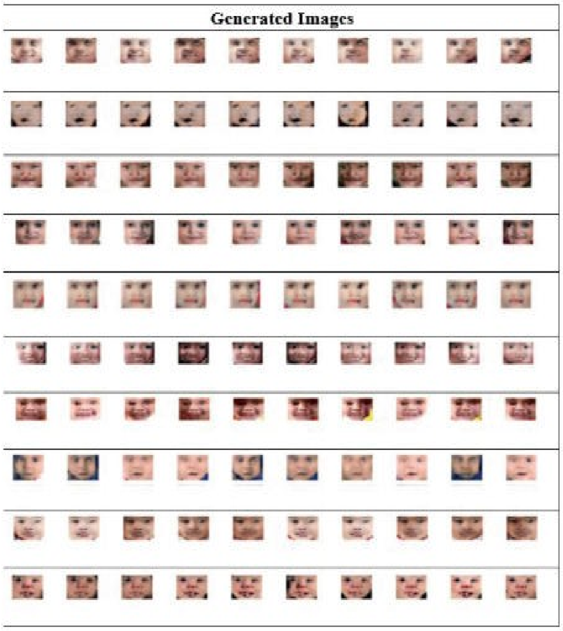


Fig. 8. Sample Output Images of the Baby Face Generation Model.

For sample output we have shown two figures. One is the Fig. 7 which is the generated output of our proposed model and another is the Fig. 8 which is the generated output of the baby face generation paper. At first glance we may think that both of these papers are very different. However, if we think clearly that both of those images are generated by a computer. And the objectives of those image creations are the same, which means creating fake human faces. If we see the baby face generation model images, they clearly lack the uniqueness of their faces. These images look like they were created with some other faces. It’s more like editing rather than creating or generating fake faces. On the other hand if we see our generated images we can clearly see that these images are not like those. And surely we can’t figure out if these images are created by mixing some other faces. It’s more like generating from the ground. Other than that we can clearly see that our image has more sharpness in eyes, nose and faces. But we have drawbacks in color and hair.

In the Table I, we can see that the different aspects of two models. In our model we have used black and white data with three different sizes of images 48*48, 100*100 and lastly 150*150. On the other hand, Baby face generation paper uses RGB images with 200*200 sizes. They ran only 200 to 300 epochs whereas our model ran about 100000 epochs. In our dataset we use 10k to 203k images on the other hand baby faces use only 623 images.

Now we can clearly see that though the sizes of images are smaller, we have used a lot of images with more epochs without the color of the image which can be trained in comparatively low computational power. Though we need more computer power than the baby face generator paper, using this extra computational power we can with proper scaling of the images create better images with sharper faces.

TABLE I. COMPARISON TABLE

Model	Image Type	Image Size	Number of Epochs	Da-taset	Size of Dataset used	Generator Loss	Output
Our Model	Black & white	48*48, 100*100, 150*150	100000	CelebA	230K, 50K, 10K	0.0-10.0	Overall better quality image, sharp Face
Baby Face Generator Model	RGB	200*200	200-300	UTKFace	623	1.0-2.0	Low quality image, RGB image

So after comparing with the baby face generation paper, we can say that in our paper using some extra computation power with proper image scaling we generated much better human faces which are more realistic and sharper. But we also have some drawbacks and our main drawback is that we have used comparatively more computational power than the baby face generation paper and our resulting images are black and white and our model needs comparatively more training data.

G. Discussion

In this paper, we try to create fake human faces. Our main focus was to create as much as a good image with low computational powers without any external image classifier. As we discussed above we used two sequential layers and a CelebA dataset to create fake human faces.

From our result and the loss function, we can see that, we successfully created human faces from random noises using limited computational power. However, the images are not as good as the real life images, so there is lots of scope in our paper to improve. However, we can see from the result that we have successfully created fake human faces from the random noise data. This is one of our main focuses.

Overall, using our low computational power we have tried to generate real life fake human faces. For comparing this model's results and to see how it reacts with different sizes of images we have tested it with three different sizes. At first, we tested it using 48 * 48 sizes. At that time, the model did not give that much good result though it generated fake human faces having noises. And then, to improve the quality we used 100*100 sizes. This time, the model performed quite well, where it gave quite a good result. When we used 100*100 sizes of the generated images were very clear and understandable. To get a better result, we then used 150*150 images. This time the model gave comparatively better results than the previous two models. Each time we ran our model at 100000 epochs.

V. CONCLUSION

In this work, a generative adversarial network model is proposed to generate fake human faces with low computational cost. The proposed generative adversarial network model generates comparatively better real life fake human faces with respect to the present state-of-the-art. We have also shown how we can only use two fully connected sequential models, one as a generator and another one as a discriminator to generate fake human faces. Though there are a lot of improvement areas in our proposed method, we can say that using our proposed model we have successfully produced pretty good fake human faces comparatively using very low computation power. In future, we can develop model to generate the images that can be used by police to track down people who are involved in adultery. We can also develop models for generating fake computer generated images in visual art, for the advertising industries and computer games.

REFERENCES

[1] Z. Zhang, X. Pan, X., S. Jiang, & P. Zhao, "High-quality face image generation based on generative adversarial networks," Journal of Visual Communication and Image Representation, Vol. 71, 2020.

[2] X. Wang, & A. Gupta, "Generative image modeling us- ing style and structure adversarial networks," In the European conference on computer vision, pp. 318-335, Springer, Cham, 2016.

[3] F. S.Ghatas, & , E. E. Hemayed, "Gankin: generating kin faces using disentangled gan," SN Applied Sciences, Vol. 2, No. 2, pp. 1-10, 2020.

[4] A.Hamdi, & B.Ghanem, " IAN: Combining Generative Ad- versarial Networks for Imaginative Face Generation," arXiv preprint arXiv:1904.07916, 2019.

[5] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, & J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," Information Fusion, Vol. 64, pp. 131-148, 2020.

[6] T. Karras, S. Laine, & T. Aila, "A style-based generator architecture for generative adversarial networks," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4401-4410, 2019.

[7] J. Zhao, X. Xie, L. Wang, M. Cao, & M. Zhang, "Generating photographic faces from the sketch guided by attribute using GAN," IEEE Access, Vol. 7, pp. 23844-23851, 2019.

[8] S. N. Esfahani, & S. Latifi, "Image generation with gans- based techniques: a survey," AIRCC's International Journal of Computer Science and Information Technology, Vol. 11, No. 5, pp. 33-50, 2019.

[9] G. Ortaç, Z. Doğan, Z. Orman, & R. ŞAMLI, " Baby Face Generation with Generative Adversarial Neural Networks: A Case Study," Acta Infologica, Vol. 4, No.1, pp. 1-9, 2020.

[10] H. Huang, P. S. Yu, & C. Wang, "An introduction to image syn- thesis with generative adversarial nets," arXiv preprint arXiv:1803.04469, 2018.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, ... & Y. Bengio, "Generative adversarial nets," Advances in neural information processing systems, Vol. 27, 2014.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, ... & Y. Bengio, "Generative adversarial networks," Communications of the ACM, Vol. 63, No. 11, pp. 139-144, 2020.

[13] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, & A. A. Bharath, "Generative adversarial networks: An overview," IEEE Signal Processing Magazine, Vol. 35, No. 1, pp. 53-65, 2018.

[14] X. Liu, B. V. Kumar, Y. Ge, C. Yang, J. You, & P. Jia, "Normalised face image generation with perceptron generative adversarial networks," In 2018 IEEE 4th International Conference on Identity, Security, and Behaviour Analysis (ISBA), pp. 1-8, IEEE, 2018.

Towards Adapting Metamodelling Technique for an Online Social Networks Forensic Investigation (OSNFI) Domain

Aliyu Musa Bade[1]
Department of Computer Science, Yobe State University
Damaturu, Nigeria

Siti Hajar Othman[2]
School of Computing, Universiti Teknologi Malaysia
Johor Bahru, Malaysia

***Abstract*—With the ease of use of smart devices, most data is now kept and exchanged in digital forms such as images, diaries, calendars, movies, and so on. Digital forensic investigation is a new technology that emerged from criminals' who extensively use computers and digital storage devices to commit different types of crimes. To address this issue, a new domain called Online Social Networks Forensic (OSNF) was created to investigate these dynamic crimes perpetrated on social media platforms. OSNFI seeks to obtain, organise, investigate, and visualise user information as direct, objective, and fair evidence. Considering the millions of individuals using social media to share and communicate, they are becoming increasingly relevant for criminal investigations. In forensics investigation of online social network, there are currently major problems such as: lack of structured procedures, the lack of unified automated methods, and the lack of a theoretical context. The use of non-uniform and ad hoc forensic techniques and procedures not only reduces the effectiveness of the process, but also affects the reliability and creditability of the proof in criminal proceedings. As a result, this paper will provide a method derived from the software engineering domain known as metamodelling, which will integrate OSNFI knowledge into an artifact known as a metamodel.**

***Keywords*—Online social networks forensic; online social networks forensic investigation; metamodelling; metamodel; model**

I. Introduction

Most organisations are now heavily reliant on digital media for information storage as digital media is used to create, process, store, and share the majority of information. Computer crimes and frauds are on the rise as the number of people using digital media grows and the law enforcement agencies faced several challenges as a result of the growing fraud and security threats. OSNs have received a lot of forensics investigation due to their widespread use and availability of supporting application interfaces. Furthermore, the number of criminal acts involving OSNs is increasing on a daily basis. It is vital to comprehend and evaluate online social network crimes and attacks in order to avoid illegal activity, discover malevolent users, and solve criminal cases. Furthermore, OSN users' safety should be enhanced to the greatest extent possible.

There are a variety of digital forensic investigation models available. However, most of them use similar methodologies and ignore the essential differences and special needs of online social networks. The main challenge is the use of a non-uniform and ad-hoc forensic techniques in the existing DFIM's [1]. The employment of non-standard and ad hoc forensic techniques and processes not only diminishes the effectiveness of the process, but it also undermines the dependability and creditability of the evidence in criminal cases.

Furthermore, while most of the created models focused more on a certain platform or content than the entire OSN [2], some still required manual treatment and could reduce the dependability and trustworthiness of evidence in criminal proceedings [3].

Therefore, there is the need to create a metamodel (a conceptual framework) for a domain which will assist many newcomers and stakeholders of the domain to have clear understanding and views of the relationship between concepts in the domain. Therefore, this is where this research will contribute in unifying the knowledge and all the processes of the OSNFI domain.

The remaining sections of this paper are organized as follows. Section II discusses on the concept of OSNFI. The OSNFI issues and challenges are presented in Section III. Section IV discusses the metamodelling approach. The preliminary results of the OSNFI metamodel development steps are presented in Section V. Sections VI and VII contain the conclusion and future work respectively.

II. Online Social Networks Forensic Investigation

The Scientific Working Group (SWG) defines Digital Forensic as the application of systematically developed and established methodologies for the safeguarding, gathering, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence produced from digital sources [4]; to ensure the correct presentation of crime evidentiary data in court [5], primarily by protecting the data's credibility and ensuring a strict chain of custody. The ultimate aim of digital forensics is to gather evidence in order to answer questions like: What transpired, Who was involved, When did it take place, Where did it take place, Why did that happen, and How an incident occurred [6]. According to [7], the prevalence of computers and other devices has caused a proliferation in the quantity of incidents and amount of digital information in society. Previously, digital forensics was primarily handled by government agencies, but has recently become more frequent in the commercial sector.

According to [5], there are many reasons to justify why digital forensic investigation models need to be developed. Among which are: *(*i). the avoidance of future malicious events against the intended target, (ii). The successful tracing of the circumstances which lead to the crime and the identification of the parties involved, (iii). Identifying and apprehending the culprits of the crime, (iv). Improving present prevention procedures in place to avoid such an event from happening again, (v). Improving corporate security experts' usage of standards to secure their respective corporate networks, and (vi). How everyone connected into this digital world can become more aware of current vulnerabilities and preventive measures. However, multiplicity, measure cloud resources, secure of digital evidence, confidentiality, hiring, training, and development are some operational challenges faced by digital forensic [4].

Forensics is utilized on social networking networks such as Facebook, MySpace, Twitter, and LinkedIn. It is well known as social media forensics, and it's a subset of digital forensics and network forensics [8]. OSNs are web-based services which allow users to create a public or semi-public profile within a limited system [3], articulate a list of other individuals that they share a link, display and navigate their list of connections as well as those created within the system by others [9]; [10]. Different SNSs, like Facebook, Twitter, and LinkedIn, are used to connect people and enable them to communicate with one another [11]. People build personal profiles from various social networking sites to share their thoughts, photographs, images, emails, and instant messaging [12], as well as to find old friends or people with common interests or problems through various social networking sites [13].

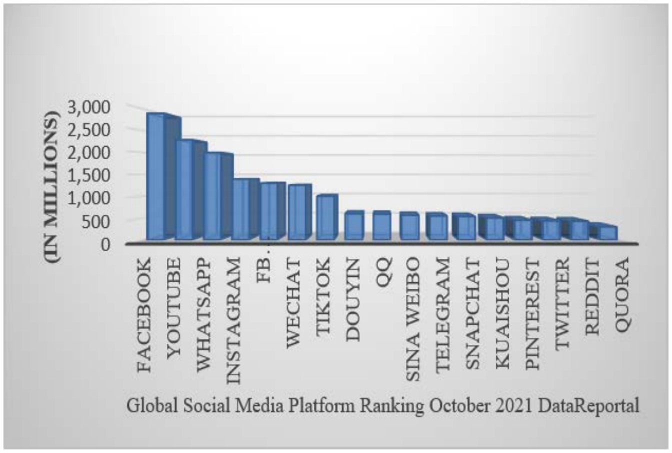


Fig. 1. Global Social Media Platform Ranking October 2021 DataReportal [14].

Facebook is managing 2,895 million users around the globe, whatsApp managing 2,000 million users, instagram managing 1,393 million users, and twitter managing 436 million users as of October 2021 as shown in Fig. 1. With the proliferation of mobile phones, the use of social network services (SNS) has skyrocketed, this SNS stores a variety of data, including user conversations, user location information, personal networks, and user psychology which can be valuable evidence in a digital forensics investigation of an incident [15]. Other uses of social networking sites include, general chatting, broadcasting breaking news, setting up a date, tracking election results, planning disaster response, humour, and serious analysis [12].

III. OSNFI Issues and Challenges

Investigators and legal practitioners are currently finding it difficult to examine social networking sites for evidence [16]. In order to discuss on the challenges associated to this explicitly, the method of the investigation are categorized into two: Conventional and Automated/Semi-automated OSN digital forensic investigation.

A. Conventional OSN Digital Forensic Investigation

Conventional OSN digital forensic investigation is a manual way of collecting and analysing all relevant information which can be considered as evidence. The following are some of the challenges associated with this technique:

1) Traditional digital forensic procedures include seizing and gathering everything that may be considered evidence. Nevertheless, this level of complete data extraction and preservation cannot be possible on online social network because of the extremely dispersed nature of social networks, their enormous scale, and mutual proprietorship of data.

2) Gathering data from people connected to the subjects (i.e., suspects or victims) lacking a reasonable suspicion of wrongdoing is virtually impossible and legally prohibited under confidentiality acts.

3) The evidence-collection process on social media sites is a process that is sometimes iterative. Therefore, the investigators will be required to gather more information if the examination subsequently identified other suspect [17].

B. Automated / Semi-automated OSN Digital Forensic Investigation

1) It can dumb down the profession because expert expertise cannot be derived solely from the field; it must also be derived from ongoing formal and informal research [13].

2) Digital forensics automation is more than just a technological issue; it also has legal and even political implications. Therefore, automation should be used only in specific phases and under expert supervision [13].

3) There are limited available automated tools that can be used in the investigation as a result of the heterogeneity of online social network and the non-existence of standardization [17].

4) Digital investigation training, practitioner training was a major concern, with 73 percent of respondents believing they don't get enough, especially in digital forensics, online investigations, and computer and network security as contained in a report by in a report by the High Technology Crime Investigation Association (HTCIA).

C. Current Issues of OSN Digital Forensic Investigation

Five relevant current issues of OSN forensic investigation as presented in Fig. 2 are discussed in this section as follows:

1) Increase in cyber-criminal activity in OSN: When creating a profile on many OSN sites, users can include their complete name, pictures, date of birth, up-to-date location, contact numbers, residential address, and office address (amongst others) [3]. Such profiles can aid in the connection of users. Users may choose to keep their profiles private, limiting their contacts, or make them public, allowing everyone to access and contact them.

Criminals may use all of the details available in these profiles to identify a specific person [18]. By their very nature, social networks have certain inherent properties that make them suitable for an adversary to manipulate. The following are the most critical of these characteristics: (i) a huge and extremely distributed user- base, (ii) clusters of users having the same social interests, developing trust with one another, and pursuing access to the same resources, and (iii) platform openness for deploying fraud resources and applications that trap users to install them.

Undoubtedly, these characteristics are the causes why cyber criminals believe there is a large chance to use online social network as a platform to commit crimes [3]. Among other illegal activities that have become a significant threat to OSN includes; Website phishing, Online sexual predators, OSN as vehicles for reaching an international audience, soliciting funding, recruiting new members, and disseminating propaganda.

2) Anti-Forensic techniques: Technical difficulties make obtaining data from the OSNs difficult. Criminals use anti-forensic techniques to hide evidence or to distract attention away from the investigation. Encryption, steganography, covert channelling, storage space data hiding, and residual data wiping are all methods used to conceal evidence [19].

3) Standard models: A scarcity of standardisation, as well as a theoretical structure for the field of digital forensics, is one of the most serious issues that investigators are currently facing [20]. A such, the use of ad hoc methods and tools for eliciting digital evidence may limit the evidence's reliability and legitimacy, particularly in criminal cases where both the evidence and the processes used to gather it can be contested [3].

4) Legal challenges: The complications in investigations as a result in scarcity of commonly established rules and criteria governing the field are referred to as legal challenges. There is no unified legal system that applies to all jurisdictions. Despite the fact that internet use has risen globally. Many countries formulate policies in accordance with their regulatory structure, which differs from country to country [19].

5) Resource challenges: Advances in cybercrime, strategies in propagating them and evading investigations are made possible by technical development. The volume of data accessible for assessments for OSNs is normally enormous. The DFIs must classify the most relevant data without jeopardising the evidence's consistency. As a result, there is a pressing need for the development of new technologies and mechanisms for fighting cybercrime, as well as qualified personnel to carry them out [19].

D. Crimes Involving OSNs

Online Social Network is a database of information that criminals can use to commit various forms of crimes such as malware distribution, fraud, harassment, grooming, Assault, burglary, domestic violence, kidnapping and so on [21]. Therefore, OSN-related crimes are divided into two categories: Classical crimes and Digital crimes as presented in Fig. 3.

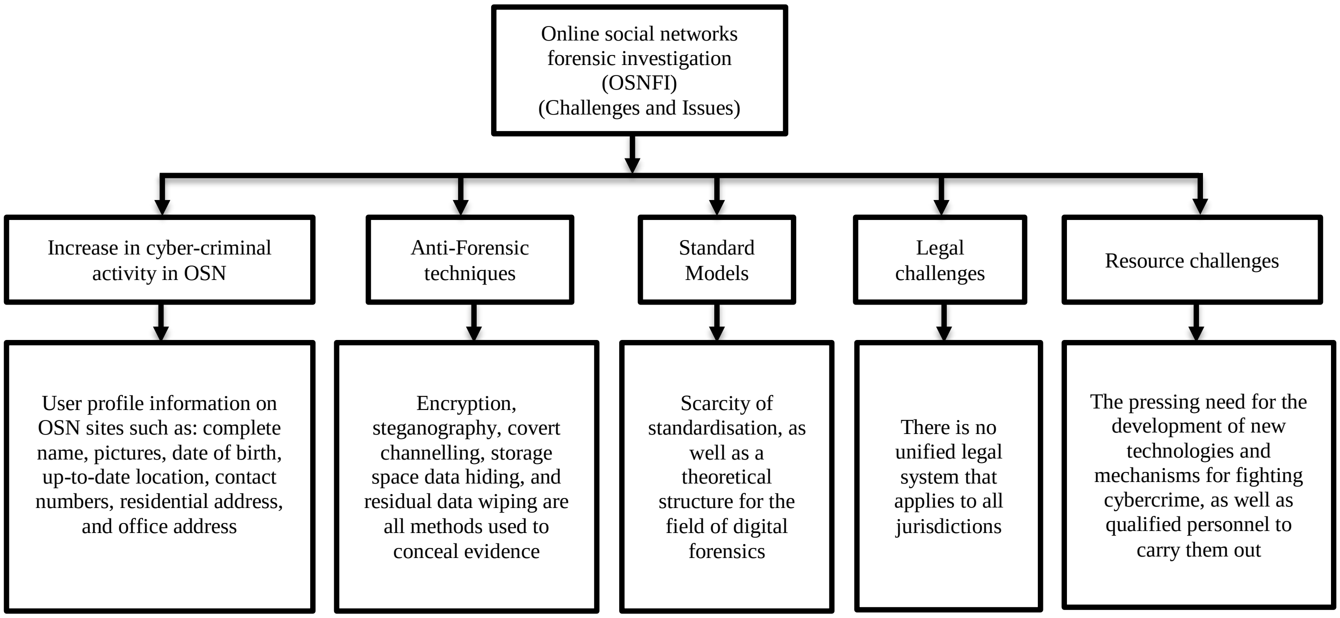

Online social networks forensic investigation (OSNFI) (Challenges and Issues)
Increase in cyber-criminal activity in OSN
Anti-Forensic techniques
Standard Models
Legal challenges
Resource challenges
User profile information on OSN sites such as: complete name, pictures, date of birth, up-to-date location, contact numbers, residential address, and office address
Encryption, steganography, covert channelling, storage space data hiding, and residual data wiping are all methods used to conceal evidence
Scarcity of standardisation, as well as a theoretical structure for the field of digital forensics
There is no unified legal system that applies to all jurisdictions
The pressing need for the development of new technologies and mechanisms for fighting cybercrime, as well as qualified personnel to carry them out

Fig. 2. OSNFI Challenges and Issues.

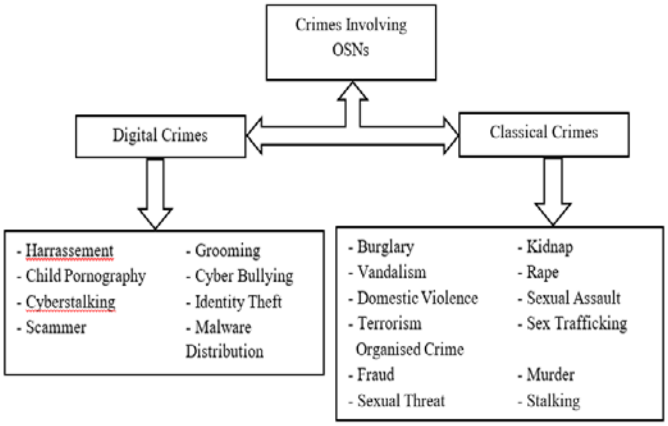


Fig. 3. Crimes Involved in OSNs [21].

1) *Classical crimes:* Online social network is referred to as a location where criminals can commit traditional criminalities. For example, social media users can update their status by posting their current whereabouts, the time they will be absent from home, and what they will be doing, giving possible criminals enough time to break into their home. This is one of several incidents that have been reported in the media [21]. Other crimes may include: vandalism, domestic violence, terrorism organised crime, fraud, sexual threat, kidnapping, rape sexual assault, sex trafficking, murder and stalking.

2) *Digital crimes:* Any criminal activity involving an information technology infrastructure, such as unauthorised or illegal entry, surveillance, and so on, is referred to as digital crime. The most common OSNs digital crimes are: cyber-based, and social engineering which is one tool that can be used to commit these crimes [21]. Other crimes may include: drug network, street gang, bombing, harassment, child phonography, cyberstalking, scammer, grooming, cyberbullying, identity theft, and malware distribution as presented in Fig. 4.

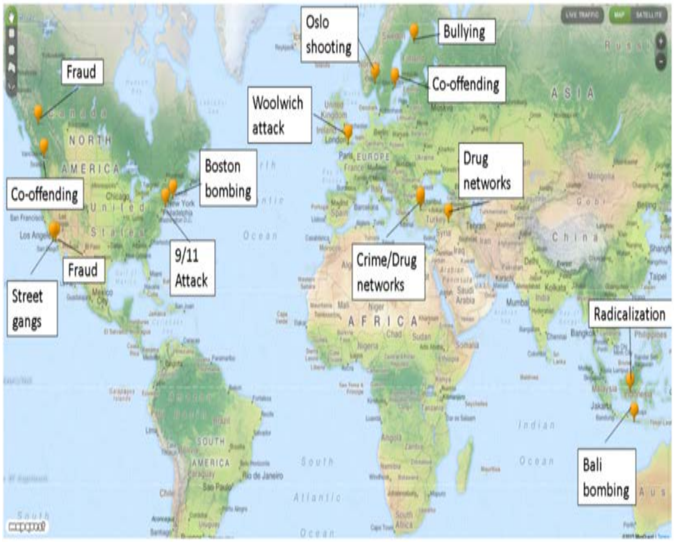


Fig. 4. Global Map of Sample Crime Analysis [22].

E. *Necessity for OSN Digital Forensic Investigation Metamodel*

SNS stores a variety of information about its users which in an investigation case, the suspect's personal data stored on a social media site will support in a variety of ways, including identifying living habits, determining geographical location, and assessing ideas and mental state. *However*, by simply collecting data from an SNS user's personal information section, it is difficult to gather useful data for a case investigation [15]. In order to address these flaws, OSN requires the establishment of a standardised forensic investigation procedure which can help investigators in an investigation. These procedures are presented in Fig. 5.

IV. Metamodelling Approach for Online Social Networks Forensic

Metamodelling is the process of developing metamodels [23]. Modelling and metamodelling are both talking about model creation but the only difference between them is in how they are interpreted.

There are numerous digital forensic investigation models available. However, the majority of them employ comparable approaches and disregard the crucial distinctions and unique requirements of online social networks.

In [17], proposed a seven phase semi-automated forensic investigation model for online social networks. These phases are: Pre-Investigation, Incident specification, Extraction, Preservation, Analysis, Iteration and Presentation. A WhatsApp Messenger Smartphone Forensic Investigation Analysis Against Web-Based WhatsApp was proposed by [24] which comprises of six phases, namely: Identification, Preservation, Collection, Examination, Analysis and Presentation. A four-phase Framework Analysis of IDFIF V2 in WhatsApp Investigation Process on Android Smartphones was proposed by [25]. The phases are: Preparation, Incidence response, Laboratorium process and Presentation.

A Framework for the ForensicAnalysis of User Interaction with Social Media was proposed by [26] which has four phases: Acquisition, Triage, Analysis and Presentation. A four-phase Digital Forensic Investigation Model for Online Social Networking was proposed by [21] which has: Preliminary, Investigation, Analysis and Evaluation phases.

The fundamental issue is the employment of ad hoc and non-uniform forensic techniques in the current DFIMs. The use of ad hoc and non-standard forensic procedures reduces not only the efficiency of the procedure but also the dependability and credibility of the evidence in criminal proceedings. Consequently, this is where the research will help to standardise the OSNFI domain's operations.

A. *Models and Metamodels*

A model is a representation of real-world phenomena [27]. It is a description of something [28], which are used to reason about a problem domain and design a solution in its domain [29]. Models are essential for comprehending and disseminating information about complex systems [30]; [31]. They have been and continue to be crucial in many scientific contexts as a large branch of philosophy of science is centered on models.

Models can be used for different purposes. They can come in form of graphical, mathematical or textual. Models can be used for descriptive, prescriptive or for defining the method by which a system will be implemented. Descriptive is simply describing a system's or a context's reality while Prescriptive is to determine the extent and depth of a problem's investigation [32]. On a philosophical level, one can concur that "everything is a model," because nothing can be managed by human mind unless it is "modelled." As a result, it's not astonishing that models have become crucial in technical fields like mechanics, civil engineering, and, eventually, computer science and engineering [33]. According to [34], models can be categorised into; conceptual versus data models, viewable and executable models, active and passive models, static and dynamic models in the study of information systems. Identifying concepts terminologies, meaning, definition and interconnections are some of the challenges encountered in model development; this is because concepts terminologies are too inconsistent and too ambiguous.

Metamodel is a model that describes/prescribes models [35]; [23]; [32]. It is a collection of *concepts* and *relations* that define the syntax of a model, and they are described using a model description language [36]. Metamodel is an abstraction that highlights properties of the model itself just as how a model is an abstraction of real-world phenomena [27]. It describes a collection of concepts and their relevant relationships, and it used as an abstraction filter in a specific modelling activity [37]. A modelling language is used in the creation of a metamodel and this language is termed as Metamodeling language [38].

In Model-Driven Engineering, metamodels are widely used to specify available model elements and structures. Nevertheless, metamodels are likely to evolve during development for a variety of reasons, such as changing requirements or evolving domain expertise [39]. Metamodeling is among the significant components of MDD; its main importance include the creation of a modelling language that will be utilized to accurately define metamodels [40] and is used to ensure the consistency of models during transformation [41]. Metamodelling is the study of processes, development of frames, production rules, constraints, models, and theories that can be used to extended models of intelligent software and information systems [23]. The basic goal of metamodeling is to represent data in such a way that it becomes self-contained and allows for the extension and alteration of its structure. Metamodeling must adhere to its own set of rules, which are as follows: Suitability, Completeness, Dynamic, Openness, Compatibility, Consistency, Reusability, and Simplicity.

B. *Step-by-Step of OSNFI Metamodel Development.*

A uniform representation of language is frequently required to reflect shared understanding in a consistent manner that fulfills the needs of the various parties involved. This method discovers specific domain characteristics, gathers domain concepts, and divides domain problems into sub-domain problems. The steps in the procedure are as follows:

1) Step 0: Identification of common phases of domain.
2) Step 1: Models collection and classification.
3) Step 2: Extraction of concepts.
4) Step 3: Identification of common concepts.
5) Step 4: Short-listing and reconciliation of definitions.
6) Step 5: Classification of common concepts: Concepts are assigned to one of the OSNFI phases: preliminary, acquisition/preservation, analysis, or presentation.
7) Step 6: Identification of relationships.
8) Step 7: Metamodel validation (*is not covered in this paper*).

To achieve the goal of creating a universally applicable OSNFIM, a broad coverage across concepts is required. Looking at the coverage measure alone, as shown in Table I, it gives a rapid indication of the supplied model's applicability. If a model can cover all of the stages of OSNFI, it is said to have a high coverage value. If the model just describes a single OSNFI phase, the coverage value of the model is lower.

TABLE I. OSNFI PROCESS MODELS

Models		**Coverage**	**Coverage of Model (according to phases)**
1	A Digital Forensic Investigation Model for Online Social Networking [3]	0.2	Preliminary, Analysis and Presentation
2	Online Social Networks As Supporting Evidence: A Digital Forensic Investigation Model and Its Application Design [9]	0.2	Preliminary, Analysis and Presentation
3	A Framework for the Forensic Analysis of User Interaction with Social Media [26]	0.2	Acquisition, Analysis and Presentation
4	Forensic Imaging for Online Social Networks [11]	0.2	Preliminary, Analysis and Presentation
5	A Review of Using Online Social Networks for Investigative Activities [21]	0.2	Acquisition, Analysis and Presentation
6	A comprehensive digital forensic investigation process model [42]	0.3	Preliminary, Acquisition, Analysis and Presentation
7	WhatsApp Messenger Smartphone Forensic Investigation Analysis Against Web-Based WhatsApp [24]	0.2	Acquisition, Analysis and Presentation
8	Framework Analysis of IDFIF V2 in WhatsApp Investigation Process on Android Smartphones [25]	0.2	Acquisition, Analysis and Presentation
9	Forensic investigation of cross platform massively multiplayer online games: Minecraft as a case study [2]	0.2	Preservation, Analysis and Presentation
10	A semi-automated forensic investigation model for online social networks [17]	0.3	Preliminary, Acquisition, Analysis and Presentation

V. The Resultant Online Social Networks Forensic Investigation Metamodel (OSNFIM)

In Fig. 5, the preliminary results of developing OSNFIM are represented by twenty-two key concepts that were collected from nineteen and filtered based on the four phases (Preliminary, Acquisition/Preservation, Analysis, and Presentation). Disparities between definitions are addressed in this process. All of the definitions listed in this process are taken into account when picking or synthesizing the common concept definition to be used as presented in Table II.

Although OSNFI is no longer widely used in research, the precise meaning of important terminology and phrases used in its concepts can often still differ among the few academics who are familiar with it. This can be a result of the researchers having different backgrounds and viewpoints. If two or more sources employ concept definitions in conflict, a technique to harmonize and fit the definition in the metamodel is necessary. In some circumstances, some models don't participate in the reconciliation procedure because they don't describe some of their concepts explicitly.

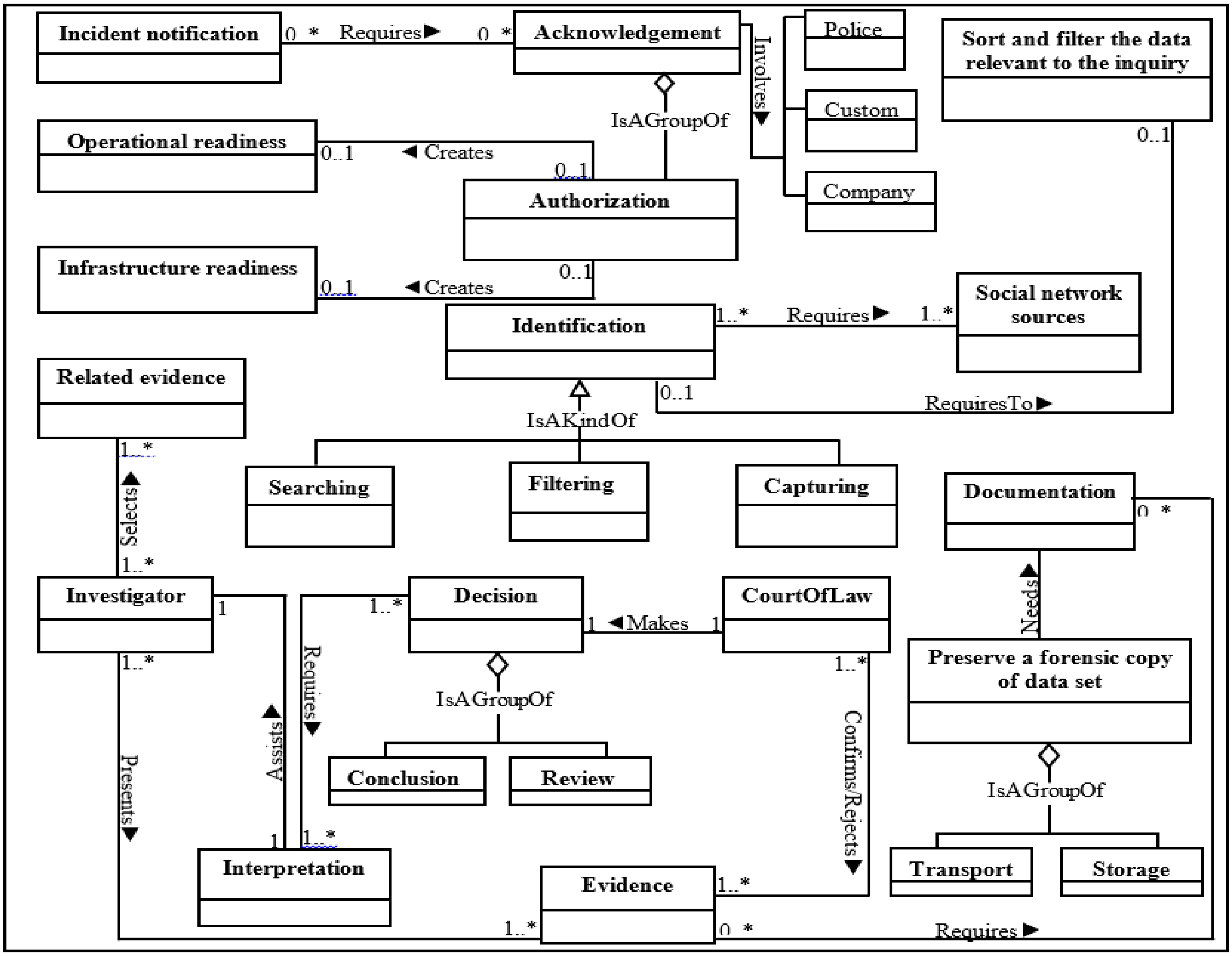


Fig. 5. The Initial Result of Online Social Networks Forensic Investigation Metamodel (OSNFIM).

TABLE II. CONCEPTS AND THEIR DEFINITIONS

S/No.	CONCEPT	DEFINITION
1	Operational readiness	This is a proper initial policy and procedure documents before a successfully launch of an effective digital forensic investigation.
2	Infrastructure readiness	Infrastructure readiness component is determined by elements external and internal to the organization which is critical to test them before implementation so that they are ready for use when needed.
3	Incident notification	Explicitly specifying the incident under investigation.
4	Authorization	Stages to gain access to evidence and legal status of the inquiry process.
5	Acknowledgment	This is the first step of an online social network forensic investigation where a case or an audit is requested from an external organisation such as the police, customs, or a company.
6	Identify Social Network sources	A way of identifying the social network involved by Initialize the SN source.
7	Identification	A process of identifying any evidence or supporting information that might be available in an online social network.
8	Searching	This is a process of discovering relevant data automatically based on the relevant data gathered from the investigation process.
9	Filtering	An activity which will scale down and focus the investigation on relevant information and discard any irrelevant information.
10	Capturing	Information collected through filtering will be captured in the best way to ensure the integrity of the data is sustained.
11	Transport	This is the process of moving digital evidence from the scene to the forensic digital laboratory.
12	Storage	A process of keeping potential digital evidence which might be needed if the analysis cannot be performed right away or if there is a legal requirement to keep digital evidence for a certain period of time.
13	Preserve a forensic copy of Data Set	Safeguarding the integrity of the original digital evidence.
14	Sort and filter the data relevant to the inquiry	The investigators must examine the results in the context of a given incident.
15	Conclusion	Investigators will conclude their examination in this stage. They may confirm their hypothesis, or they might need to find more information related to the entities involved in an investigation to credit or discredit a theory.
16	Select Relevant Evidence	The investigators will select the evidence that is relevant and appropriate for presenting in court.
17	Present the Evidence	To show anything you see, experience, read, or hear that leads you to assume something is true or has actually happened
18	Decision	A decision or resolution reached after careful consideration
19	Interpretation	To use scientifically proven methods to explain facts discovered throughout the analysis process within the context of the investigation.
20	Documentation	Documentation of all activities that have been done from the beginning of the investigation process to the end of the analysis process in the forensic laboratory.
21	Investigator	A person who conducts a formal investigation or inquiry
22	CourtOfLaw	A group of individuals presided over by a judge or judges who operate as a tribunal in civil and criminal proceedings

VI. CONCLUSION

Online social networks forensic investigation domain is a new, but extremely important and a high demand domain. The number of crimes increases on daily basis due to the advancement in technology and the use of smart devices. The problems with the OSNF domain which are addressed in this paper are: lack of uniformity in the procedures for the OSN investigation, increase in cyber-criminal activity, anti-Forensic techniques, standard models, legal and resource challenges. Hence, the OSNF investigation metamodel proposed will aid in the proper investigation of digital crimes through various processes.

VII. FUTURE WORK

The future work will be to validate the proposed Metamodel based on two relevant metamodel validation technique: (i) Comparison to other models and, (ii) Frequency-based Selection validation techniques.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Yobe State University (YSU), Damaturu, Nigeria and the Tertiary Education Trust Fund (TETFund) of Nigeria for funding this research.

REFERENCES

[1] Bade AM, Othman SH. A Systematic Review of Published Articles, Phases and Activities in an Online Social Networks Forensic Investigation Domain. Int. J. Adv. Comput. Sci. Appl. 2021;12:153–60.

[2] Taylor DCPJ, Mwiki H, Dehghantanha A, Akibini A, Kwang K, Choo R, et al. Science & Justice Forensic investigation of cross platform massively multiplayer online games : Minecraft as a case study. Sci. Justice [Internet] 2019;59:337–48. Available from: https://doi.org/10.1016/j.scijus.2019.01.005.

[3] Zainudin, M N, Merabti, Madjid, Llewellyn-jones, David. A Digital Forensic Investigation Model for Online Social Networking. 2010;1–6.

[4] Vincze EA. Challenges in digital forensics. Police Pract. Res. 2016;17:183–94.

[5] Kaur R, Kaur A. Digital Forensics. Int. J. Comput. Appl. 2012;50:5–9.

[6] Dimitriadis A, Ivezic N, Kulvatunyou B, Mavridis I. Digital forensics framework for reviewing and investigating cyber attacks. Array [Internet] 2020;5:100015. Available from: https://doi.org/10.1016/j.array.2019.100015.

[7] Walker N, Kebande VR. Conference Title : The International Conference on Digital Security and Forensics (DigitalSec2014) Conference Date : June 24-26 , 2014 Conference Venue : VSB-Technical University of Ostrava , Czech Republic ISBN : Published by : The Society of Digital I. 2014.

[8] Chang C-P. Knowledge Production from Social Network Sites - Using Social Media Evidence in the Criminal Procedure (Title of the Thesis) Knowledge Production from Social Network Sites - Using Social Media Evidence in the Criminal Procedure. 2014.

[9] Mohd Zainudin N, Merabti M, Llewellyn-Jones D. Online social networks as supporting evidence: A digital forensic investigation model and its application design. 2011 Int. Conf. Res. Innov. Inf. Syst. ICRIIS'11 2011.

[10] Power A. What is social media? 2012.

[11] Kale S, Sahu PA. Forensic Imaging for Online Social Networks. 2014;3:166–70.

[12] Montasari R. Digital Forensic Investigation of Social Media , Acquisition and Analysis of Digital Evidence. 2019;2:52–60.

[13] Kleinberg JM. Challenges in mining social network data. 2007;13:4–5.

[14] KEMP S. DIGITAL 2021 OCTOBER GLOBAL STATSHOT REPORT. Datareportal2021.

[15] Jang YJ, Kwak J. Digital forensics investigation methodology applicable for social network services. Multimed. Tools Appl. 2015;74:5029–40.

[16] Lu R, Li L. Research on forensic model of online social network. 2019 IEEE 4th Int. Conf. Cloud Comput. Big Data Anal. ICCCBDA 2019 2019;116–9.

[17] Arshad H, Omlara E, Oludare I, Aminu A. Computers & Security A semi-automated forensic investigation model for online social networks. Comput. Secur. [Internet] 2020;97:101946. Available from: https://doi.org/10.1016/j.cose.2020.101946.

[18] Athanasopoulos E, Makridakis A, Antonatos S, Antoniades D. Antisocial Networks : Turning a Social Network into a Botnet. 2008;1–15.

[19] Fakiha B. Journal of the Arab American University مجلة الجامعة العربية الامريكية للبحوث Digital Forensics : Crimes and Challenges in Online Social Networks Forensics Digital Forensics : Crimes and Challenges in Online Social Networks Forensics. 2020;6.

[20] Arshad H, Jantan A, Omolara E. Evidence collection and forensics on social networks: Research challenges and directions. Digit. Investig. [Internet] 2019;28:126–38. Available from: https://doi.org/10.1016/ j.diin.2019.02.001.

[21] Abdalla A, Yayilgan SY. A Review of Using Online Social Networks. 2014;8531:3–12.

[22] Karabiyik U, Akbas E, Canbaz MA, Aksoy A, Tuna T, Gonen B, et al. Journal of Digital Forensics , Security and Law A Survey of Social Network Forensics. 2016;11.

[23] Savchenko Y, Stepashko V. Metamodeling as a way to universalization of inductive modeling tools. 2018 IEEE 13th Int. Sci. Tech. Conf. Comput. Sci. Inf. Technol. CSIT 2018 - Proc. 2018;1:444–7.

[24] Anwar N, ImamRiadi. Forensic Investigation Analysis of WhatsAppMessenger Smartphone Against WhatsApp Messenger Smartphone Forensic Investigation Analysis Against Web-Based WhatsApp. 2017;3:1–10.

[25] Rahman D, Rahadhian, Riadi I. Framework Analysis of IDFIF V2 in WhatsApp InvestigationProcess on Android Smartphones. Int. J. Cyber-Security Digit. Forensics 2019;8:213–22.

[26] Haggerty J, Casson MC, Haggerty S, Taylor MJ. A framework for the forensic analysis of user interaction with social media. Int. J. Digit. Crime Forensics 2012;4:15–30.

[27] Tech V, Ikipedia WS. Metamodeling : What is it good for ? 2009;94085.

[28] PYLE D. What Is a Model? Bus. Model. Data Min. 2003;91–119.

[29] Brown AW, Conallen J, Tropeano D. Introduction: Models, Modeling, and Model-Driven Architecture (MDA). 2005.

[30] Link S, Hoyer P, Schuster T, Abeck S. Model-driven development of human tasks for workflows. Proc. - 3rd Int. Conf. Softw. Eng. Adv. ICSEA 2008, Incl. ENTISY 2008 Int. Work. Enterp. Inf. Syst. 2008;329–35.

[31] Rutle A, Simonsen KIF, Schaathun HG, Kirchhoff R. Model-driven software engineering in practice: A content analysis software for health reform agreements. Procedia Comput. Sci. 2015;63:545–52.

[32] Yonglin LEI, Zhi ZHU, Qun LI. An ontological metamodeling framework for semantic simulation model engineering. 2020;31:527–38.

[33] Brambilla M, Cabot J, Wimmer M. Model-Driven Software Engineering in Practice. 2017.

[34] Othman SH. Metamodelling Approach for Managing Disaster Management Knowledge. 2012.

[35] Liu J, Zhang N, Kong X, Gu Y. Research on metamodeling process of E-government affair system. Proc. - 2012 Int. Conf. Comput. Sci. Electron. Eng. ICCSEE 2012 2012;1:566–70.

[36] Trabelsi C, Ben Atitallah R, Meftali S, Dekeyser JL, Jemai A. A model-driven approach for hybrid power estimation in embedded systems design. Eurasip J. Embed. Syst. 2011;2011.

[37] Bzivin J, Nantes U De, Houssinikre D, Bezivin J, Gerb O, Montral HEC. Towards a Precise Definition of the OMGMDA Framework. 2001.

[38] Karagiannis D, Kühn H. Metamodelling Platforms. 2002.

[39] Demuth A. Enabling dynamic metamodels through constraint-driven modeling. Proc. - Int. Conf. Softw. Eng. 2012;1622–4.

[40] Liu Y, Wang Y. A study of metamodeling based on MDA. ICCRD2011 - 2011 3rd Int. Conf. Comput. Res. Dev. 2011;2:171–3.

[41] Herr S, Wirtz G. The 20 International Conference on. 2008.

[42] Mir SS, Shoaib U, Sarfraz MS. Analysis of Digital Forensic Investigation Models. 2016;14:292–301.

Recommendation System based on User Trust and Ratings

Mohamed TIMMI[1], Loubna LAAOUINA[2], Adil JEGHAL[3], Said EL GAROUANI[4], Ali YAHYAOUY[5]
LISAC Laboratory, Faculty of Sciences Dhar El Mehraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco[1, 4, 5]
LISA Laboratory, National School of Applied Science, Sidi Mohamed Ben Abdellah University, Fez, Morocco[2]
LISAC Laboratory, National School of Applied Science, Sidi Mohamed Ben Abdellah University, Fez, Morocco[3]

***Abstract*—Recommendation systems aim at providing the user with large information that will be user-friendly. They are techniques based on the individual's contribution in rating the items. The main principle of recommendation systems is that it is useful for user's sharing the same interests. Furthermore, collaborative filtering is a widely used technique for creating recommender systems, and it has been successfully applied in many programs. However, collaborative filtering faces multiple issues that affect the recommended accuracy, including data sparsity and cold start, which is caused by the lack of the user's feedback. To address these issues, a new method called "GlotMF" has been suggested to enhance the collaborative filtering method of recommendation accuracy. Trust-based social networks are also used by modelling the user's preferences and using different user's situations. The experimental results based on real data sets show that the proposed method performs better result compared to trust-based recommendation approaches, in terms of prediction accuracy.**

Keywords—Recommendation systems; collaborative filtering; trust; social networks

I. INTRODUCTION

The platforms with thousands of items will support the users to be able to know how to connect to the right content, which is relevant to their interests and concern. To help users, the systems of recommendation emerge as a great solution to personalize the content presented to the users in the form of techniques and software tools that provide personalized suggestions and recommendations for items in order to boost the users' competencies [1, 2]. Even though several types of methods have been proposed to build systems of recommendation, the collaborative filtering method remains one of the greatest widely used and adopted techniques to generate recommendations. It is far from ideal in terms of predictive performance. As, it suffers from countless inherent problems [3, 4]. The most important thing of these is data sparsity and cold start, which affects the recommender's accuracy of the system [3]. To address these issues and model the user's preferences more accurately, the additional information can be incorporated into the collaborative filtering method to compensate for insufficient rating information, such as social media information, including friendship, belonging, and trusting relationships [1, 5, 6].

The relationship which is based on trust is one of the most crucial types of social relationships, as it gives its power and good positive association with similarities between the users [1], and several studies have shown great efficiency in improving predictive accuracy compared with the traditional recommendation techniques. Additionally, collaborative filtering is one of the most common approaches in systems of recommendation. As it does not depend on additional data, only the history of interactions, it becomes quite simple to be reproduced in various real applications and increase its popularity. The recommendation based on collaborative filtering was developed from the observation that people tend to adopt other people's recommendations. Someone who has the intention to purchase a certain product, for example, s/he looks for opinions and points of view from the other people who have already purchased and bought the same product before deciding to purchase. This happens frequently in the daily lives of people with different yet varied situations. The selection of a certain movie, a book, among many other [2]. Several trust-based systems of recommendation that employ these models to solve data sparsity and cold-start problems have also been proposed to combine the impacts and the great influences of social trust with different strategies. However, the previous work which is proposed in this field failed to systematically model the reciprocal effect between the users. It cannot model how and to what extent the user's preferences are affected by trustees and at the same time to what extent it influences the same user by trustors, where the user preferences as trustee or trustor can be distinct from each other [7]. Therefore, when predicting a user's preferences for an item, it does make more sense to consider both the trustor's preferences and the trustee's performances at the same time. However, in the previous studies, the methods modelled the users using a single case [8], or by separately considering the two user cases [7]. In other words, no distinction is made between different cases of the user as trustee or trustor in the ratings generation process.

Regardless of the learning approach adopted by the systems of recommendation, there must be a past set of interactions that describe the users' relationships with the items of the system. Past interactions between a user and an item are traditionally called feedback, and they can be either explicit or implicit. Most of the existing methods depend on an explicit trust bond between the users, based on which users display their preferences as trustors or trustees, except those users who may not explicitly interact with others, but rather implicitly. We note that most of the methods found in previous studies are effective and efficient in modelling explicit relationships. However, they do not consider the discovery and modelling of implicit interactions between two users who may be similar but not connected in the network of trust. The local perspective of social relations reveals the relationship between the user and their neighbors, while the global perspective of social relations reveals the reputation of the user in the social network [9]. Users around the world are more likely to seek suggestions from their local friends. Yet, they may also be tempted to solicit suggestions from high-reputation users, indicating that public and local opinions based on social relations might be exploited to improve the performance of systems of recommendation.

As a suggestion, a model called "GlotMF" (Global Local Trust with Matrix Factorization), that exploits both the global and local social context of trust relationships for recommendations. This work introduces a new strategy for merging ratings data and trust data by sufficiently exploring how to generate known ratings under the influence of the trust behaviors of users that are intertwined in their trust network, rather than simply combining two types of data, as most previous studies did, to express the mutual influence of users more logically on each other's opinions. The proposed method uses a matrix factorization technique to model user preferences for trust-based recommendations, and the preferences of the two different cases of users are learned by modelling the explicit and implicit interactions between them. Specifically, the preferences of the trustees and trustors are estimated to be distinctly suitable for the explicit ratings and explicit trust relationships of existing methods that measure the association of two users based on only the links between them. It is also being taken advantage of the structure of Local Trust Network Links to assess links between trustees and trustors users, as the structure of these links is used to model the user's implicit interaction with other users in terms of both trustees' and trustors' preferences. Experiments conducted on a real dataset demonstrate the effectiveness of our method in terms of predictive accuracy, and the results confirm that our method achieves promising recommendation performance, especially by dint of its effectiveness for Cold-Start and sparsity data compared to its counterparts.

II. Literature Review of Related Work

The merge of model-based collaborative filtering methods with trust relationships to improve the accuracy of recommendations has recently become a very popular research topic, especially using the matrix factorization technique due to its high precision and ease of use contribution to alleviating the problem of sparsity data better than other techniques [3, 8]. Many researchers have exploited this technique to learn about latent features of users and well-known ratings items, and to merge social relationships between users with rating data using different techniques. The researchers proposed in [8] a "SoRec" model which integrates a social network database into a probability matrix factorization model by simultaneously analyzing the rating matrix and the social trust matrix by sharing the matrix of features latent to a user [10]. Their empirical analysis shows that their method is superior to the basic matrix factorization model and other memory-based methods that take advantage of trust relationships, but that true recommendation processes are not reflected in this model. Thus, to model the information of confidence in a more realistic way, the same researchers proposed in a model "RSTE"[11], which interprets the user's decision to rate like a balance between his own tastes and those of his neighbor's trustees. Their experiences shows that their model outperforms the basic matrix factorization method and existing trust-based methods, but in their model, the vectors of features of user's immediate neighbors influence his ratings rather of influence his vector of features, and this model does not deal with the diffusion of trust.

The researchers reinforced in [12] this model by allowing the diffusion of trust and built a "SocialMF" model, which integrates the social impact by making the latent features of each user depend on the latent features of their immediate neighbors in the social network. Moreover, to effectively use the information of social networks when there is no trust information available, the researchers proposed in [13, 14] a "SoReg" model that performs matrix factorization while exploiting social regularization defined based on both user-item matrices and positive social relations. This work is different from previous studies in the field of trust-based recommendation since it recognizes the difference between the relationship of trust and the relationship of friendship, as well as it forces the preferences of the user to be closer to the preferences of their friends in the social network.

The methods "SoReg", "SocialMF", "STE" and "SoRec" in general have the same goal [15, 16], while the most important work relevant to our work, which is "TrustMF", in this paper is the preferences of different cases are learned independently to guess the ratings. However, in this paper, it is said that it makes more sense to consider both the preferences of the trustee and the preferences of the trustor at the same time in the learning process since the assessment is generated from the two cases. In addition, the "TrustMF" model cannot capture the implicit relationship between the confident user and the trusted user when they are not socially related, nor does it consider the trust of the public, which will be addressed in our proposed model[17].

III. A GlotMF: A Model-based Method

A. *Problem Description*

A recommender system that involves m users and n items is introduced to introduce some notations used to model the recommendation problem in this work. Let $U = \{u_1, u_2, \ldots, u_n\}$ and $V = \{v_1, v_2, \ldots, v_m\}$ be two groups of users and items respectively, where n is the number of users and m is the number items. Let $R \in R^{nxm}$denotes the user-item rating matrix which represents the numerical scores given by the users on the items, and $R_{i,j}$ represents the rating of item v_j given by user u_i, where each user evaluates a subset item with certain values from a rating field predefined by the recommendation system. Let $\Omega = \left\{(\boldsymbol{i}, \boldsymbol{j}) \colon \boldsymbol{R_{i,j}} \neq \boldsymbol{0}\right\}$ denotes the locations of observed ratings in the rating matrix R.

$T \in R^{nxn}$ is the user's trust relationship matrix, where $T_{i,k}$ is a real number in the domain [0, 1] describing the strength of the relationship between users u_i and u_k . Let $\psi = \left\{(i, k) \colon T_{i,k} \neq 0\right\}$ denotes the locations of observed trust relations in trust network matrix T. Since we use in this paper a matrix analysis technique to build the proposed model, let $B_i \in R^k$ be the K-dimensional preference vector of the trustor and $E_i \in R^k$the K-dimensional preference vector of trustee for the useru_i. $V_i \in R^k$ is a k-dimensional feature vector of the element v_j. So, we can formulate the recommendation problem in this paper as follows: by giving a set of user ratings on R's items and a set of trust values T for users by other users who also rated a group of items, and by using the matrix analysis technique to study how to learn the preferences of the different states of the users and the features of the items to guess the rating given by the target user u_i on the target item v_j plus precisely.

B. *Matrix Factorization Model*

The matrix factorization model assumes that some latent factors influence a user's rating behaviors and that the vector of user preferences is determined by how each factor is applied to that user [18]. This hypothesis makes it possible to discover missing ratings in the rating matrix from known ratings. This technique decomposes the rating matrix R into two matrices of lower order K which are the matrix of the latent features of the user U and the matrix of the latent features of the item V that is to say $R \approx U^T V$ (as shown in Fig. 1), where the low dimension U and V matrices are unknown and must be predicted. Thus, the goal of the matrix factorization technique is to learn the matrices of latent features U and V and to use them thereafter to provide predictions of missing ratings by solving the following optimization problem [19]:

$$\min_{U,V} \sum_{i=1}^{n} \sum_{j=1}^{m} W_{i,j} \left(R_{i,j} - U_i^T V_j\right)^2 + \lambda(\|U\|_F^2 + \|V\|_F^2) \quad (1)$$

$U_i \in R^k$denotes a vector of the latent features of user's preferences u_i, and $V_j \in R^k$ denotes a vector of latent features of the preferences of the item v_j, K is the number of latent features, λ is the regulation parameter which controls the complexity of the model to avoid relevance with training data (over-fitting) by introducing the term $\lambda(\|U\|_F^2 + \|V\|_F^2)$ where $\|.\|_F^2$ This is the "Frobenius norm". Conversely, $W \in R^{nxm}$ is the weight matrix where $W_{i,j}$ is the weight of the rating given by the user u_i on the item v_j. The common way to define W is $W_{i,j} =$ 1 if $R_{i,j} \neq$ 0, but a matrix of weight W can also be used to process implicit opinions and encode secondary information such as the similarity between users and items or user reputation. This factorization is illustrated in Fig. 1.

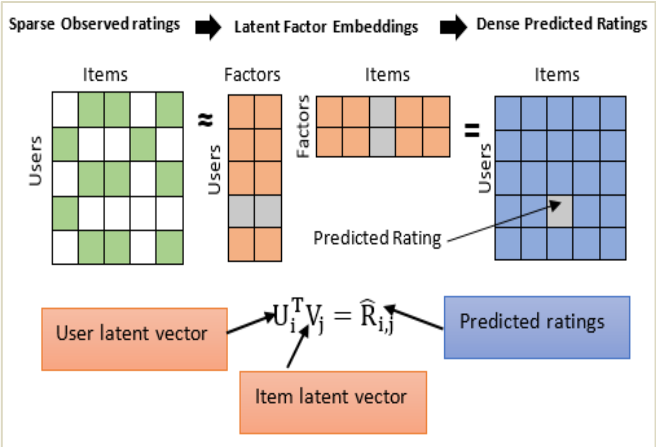


Fig. 1. Operating Principle of the Matrix Factorization Technique.

C. *Steps of the Construction: The Proposed Model*

1) Modeling of global information in social networks: The information contained in social networks represents the reputation of the user in the network, where reputation is a type of case that gives additional powers and capabilities to recommendation systems. There are many algorithms to calculate the reputation value of social network nodes based on their connections. In this work, we rely on one of the most popular algorithms, "PageRank" to calculate user reputation values. We first apply the "PageRank" algorithm to rank users by exploiting the general view of social networks, assuming that $r_i \in$ is the reputation rank of u_i so that $r_i = 1$ indicates that u_i has the highest reputation in the entire social network, then we set the reputation value w_i to u_i according to the reputation value r_i according to as follows:

$$w_i = f(r_i) = \frac{1}{1+log(r_i)} \quad (2)$$

So long as the function (f) constrains the reputation value w_i in the interval [0, 1] is a decreasing function for r_i, that is, higher rank users have high reputation values. So, to model the information in social networks, we can use the reputation values of the users to weigh the significance of their recommendations by modifying the previous equation (3) so that it becomes as follows:

$$\min_{B,E,V} \sum_{(i,j)\in\Omega} w_i \left(R_{i,j} - g(\alpha B_i^T V_j + (1-\alpha) E_i^T V_j)\right)^2 + \lambda(\|B\|_F^2 + \|E\|_F^2 + \|V\|_F^2) \quad (3)$$

During this matrix factorization, the large value of w indicating the high reputation of u_i , will force the term $(\alpha B_i^T V_j + (1-\alpha) E_i^T V_j)$ to fit tightly into the evaluation of $R_{i,j}$, while the small value ofw_i will make the term $(\alpha B_i^T V_j + (1-\alpha) E_i^T V_j)$ approximate $R_{i,j}$.

2) Modeling of local information in social networks: Local information represents preferences of the two different states of the user as a trustee and a trustor which are learned by modelling the explicit interactions between users. The preferences of the trustor and the preferences of the trustee are calculated to account for explicit ratings and explicit trust relationships. The local trust network link structure is used to assess the links between trustees and trustors, as the structure of these links is exploited as organizational boundaries to model the user's implicit interaction with other users in terms of the preferences of the trustors and the preferences of the trustees.

3) Modeling of explicit interactions between users: In this part, there is a description of how to generate ratings and trust relationships from the perspective of different user cases as shown in Fig. 2, where the social impact of user ratings can flow in both ways. That is to say, the user's rating is not influenced only by trustees, but also by trustors and this is what is confirmed by the researchers in [20], which indicates that the influence of trustors in predicting rating may be equal to that of trustees, and therefore may provide added value to predict ratings more accurately.

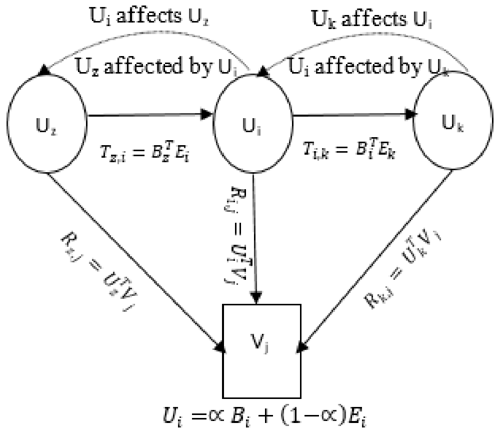


Fig. 2. The Influence of Trusters and Trustees on the Prediction of the Rating for the Target user on the Target Item.

Fig. 2 shows the proposed trustor model that can characterize how a user U_i's ratings are affected by other users they trust by means of $B_i^T V_j$, and as it does show the proposed trustee model that is able to characterize how a user U_i's opinions affect the decisions of others who trust U_i by means of $E_i^T V_j$.

a) Trust modeling: In addition to the rating data, there is a huge user-generated trust network also available on product showcases on the Internet. Following user preferences are affected by the preferences of the trustees through browsing activities and comments (For example, presenting ratings and reviews of users who trust products) and affect the trustors themselves. Along with the relationship between user preferences and trust, we can also model user preferences on the known trust database using a matrix factorization technique as shown in [18, 21]. When users are mapped to the same k-dimensional vector space, we can model each already known trust value $T_{i,k}$ as the internal product of B_i and E_k, i.e., the expected strength of the trust relationship between u_i and u_k is given by:

$$\hat{T}_{i,k} = B_i^T E_k \tag{4}$$

Considering only the trust data, we can learn the feature matrices $B \in R^{k \times n}$ and $E \in R^{k \times n}$ by solving the following optimization problem:

$$\min_{B,E} \sum_{(i,k) \in \psi} \left(T_{i,k} - g\left(\hat{T}_{i,k}\right) \right)^2 + \lambda (\|B\|_F^2 + \|E\|_F^2) \tag{5}$$

Thus, by performing the above matrix factorization, anyone can learn the user's preferences in terms of the latent features vectors of the trustee, and the latent features vectors of the trustor from the known trust data.

b) Rating modeling: Given the assumption that different user cases influence rating generation differently [6, 20], each known rating should be determined by the preferences of the trustor as well as the preferences of the trustee. Based on this, the rating is predicted by user u_i on the v_i item as follows:

$$\hat{R}_{i,j} = \propto B_i^T V_j + (1 - \propto) E_i^T V_j \tag{6}$$

Given α is the parameter to control the contribution to the evaluation of the different cases of the user. So, by considering only the rating data, we can learn the feature matrices $B \in R^{k \times n}$, $E \in R^{k \times n}$, $V \in R^{k \times m}$ by solving the following optimization problem:

$$\min_{B,E,V} \sum_{(i,j) \in \Omega} \left(R_{i,j} - g\left(\hat{R}_{i,j}\right) \right)^2 + \lambda F \tag{7}$$

where $\mathrm{F} = (\|B\|_F^2 + \|E\|_F^2 + \|V\|_F^2)$ This is the "Frobenius norm".

Given that g(x) is the logistic function suggested by the researchers to link the internal product of feature vectors latent in the interval [0, 1] is given by the relation $g(x) = 1/(1 + e^{-X})$ To learn the parameters more conveniently, we map the raw rating $R_{i,j}$ to the interval [0, 1] using the function $(x) = 1/(1 + e^{-X})$, where Rmax is the maximum value of the ratings defined in the recommender system. After training the model and learning the feature matrices, the prediction of the rating can be obtained from user u_i at item v_i by the relation $g\left(\hat{R}_{i,j}\right)$ x R_{max} Thus, by performing the previous matrix factorization technique, a person can learn the user's preferences in terms of their latent feature vectors in both cases as trustee and trustor from already known rating data.

Modeling of implicit interactions between users: In this section, we describe how to model the implicit interactions between two trustees and between two trustors by incorporating the local binding structure of the trust network to restrict the objective function.

c) The implicit influence of the trusters: Two trustors are alike if they share many out-links in a trusted network. In other words, they are jointly chatting with several trustees. So, by taking all the user's trust links with other users, and instead of relying on a single link, we can achieve a more precise and robust link between two trustors even if they are not explicitly linked. Therefore, to capture the similarity between two trustors users u_i, u_k depending on the structure of the links coming out of it, we adopt the following "cosine similarity" scale[22]:

$$S_{i,k}^B = \frac{\sum_{f=1}^n T_{i,f} T_{k,f}}{\sqrt{(\sum_{f=1}^n T_{i,f}^2) \ x \ (\sum_{f=1}^n T_{k,f}^2)}} \tag{8}$$

In our experiment, we have used a binary value for the trust $T_{i,k}$ (0 or 1), and here B denotes the similarity between two trustors. With this similarity, we can model the implicit effect of the trustors by minimizing the following term:

$$\sum_{i=1}^n \sum_{k=1}^n S_{i,k}^B \|B_i - B_k\|_F^2 \tag{9}$$

The large value of $S_{i,k}^B$ indicates that the trustor u_i and the trustor u_k share several external bonds, and we, therefore, enforce their preference vectors to be as close as possible, while a small value of $S_{i,k}^B$ indicates that the distance between the two preference vectors must be large. Therefore, by presenting this structure-dependent analogy, the vectors of the preferences of the trustors are linked in the learning process.

d) The implicit impact of trustees: Two trustees are alike if they share many in-links in the trust network. That is, they are jointly trusted by many trustors. Thus, the similarity between two trustees u_i and u_k based on the structure of the links entering them can be captured by the following scale:

$$S_{i,k}^{E} = \frac{\sum_{f=1}^{n} T_{f,i} . T_{f,k}}{\sqrt{(\sum_{f=1}^{n} T_{f,i}^{2})\ x\ (\sum_{f=1}^{n} T_{f,k}^{2})}} \tag{10}$$

Given that E here denotes the similarity between two trustees, as in modelling the implicit influence of the trustors, we model the implicit influence of the trustees by minimizing the following term:

$$\sum_{i=1}^{n} \sum_{k=1}^{n} S_{i,k}^{E} \, \|E_i - E_k\|_F^2 \tag{11}$$

4) The proposed model standardized framework: As shown above, the explicit interactions and the implicit interactions between the trustors and the trustees. Moreover, we have presented how to model the information represented by the reputation of the users in the trust networks. We then propose the following merged model which considers all the previous information and find a solution for the following objective function:

$$\mathcal{L} = \frac{1}{2}\Big(\sum_{(i,j)\in\Omega} w_i \left(R_{i,j} - g(\hat{R}_{i,j})\right)^2 + \sum_{(i,k)\in\psi} \left(T_{i,k} - g(\hat{T}_{i,k})\right)^2 + \lambda_B \sum_{i=1}^{n} \sum_{k=1}^{n} S_{i,k}^{B} \|B_i - B_k\|_F^2 + \lambda_E \sum_{i=1}^{n} \sum_{k=1}^{n} S_{i,k}^{E} \|E_i - E_k\|_F^2 + \lambda \mathrm{F}\Big) \tag{12}$$

λ_B and λ_E are respectively parameters to control the extent of the influence of the implicit interactions between the trustors and those between the trustees. To reduce the complexity of the model, we experimented with λB = λE. The former term of the previous equation is the regulation parameter used to avoid large adaptation to the training data, while λ is the regulation parameter.

D. *Learning the Model*

To get the minimum term of the previous objective function, and, thus, learn the feature matrices V, E and B and use them to predict unknown ratings, we use the Stochastic Gradient Descent method, which generally works efficiently for recommender systems.

Fig. 3 shows instructions of the algorithm for learning the model. There are several parameters taken as input, including the rating matrix R, the trust matrix T, the regulation parameters E, B, and A, the parameter controlling the rating contribution of the different cases of the user ∝, the initial learning rate μ, and the number of latent factors K.

In the first place, the researchers randomly generate the latent feature matrices V, E and B with small values. Secondly, we continue to train the model until the objective function converges to L. More precisely, we calculate the derivatives of the variables V, E and B then we modify them using the Stochastic Gradient Descent method. Finally, we obtain a latent feature vector of the trustor and a latent feature vector of the trustee and the latent feature vector of the item, and we use them to compute the prediction of the target user on the target item.

1) A learning algorithm for the proposed model

R : Rating matrix,

T : social matrix,

$\lambda, \lambda_B, \lambda_E, \alpha$ and K: hype-parameters

μ : learning rate

B and E: Features matrices for users with roles of truster and trustee,

V: Feature matrix for items with implicit and explicit feedbacks.

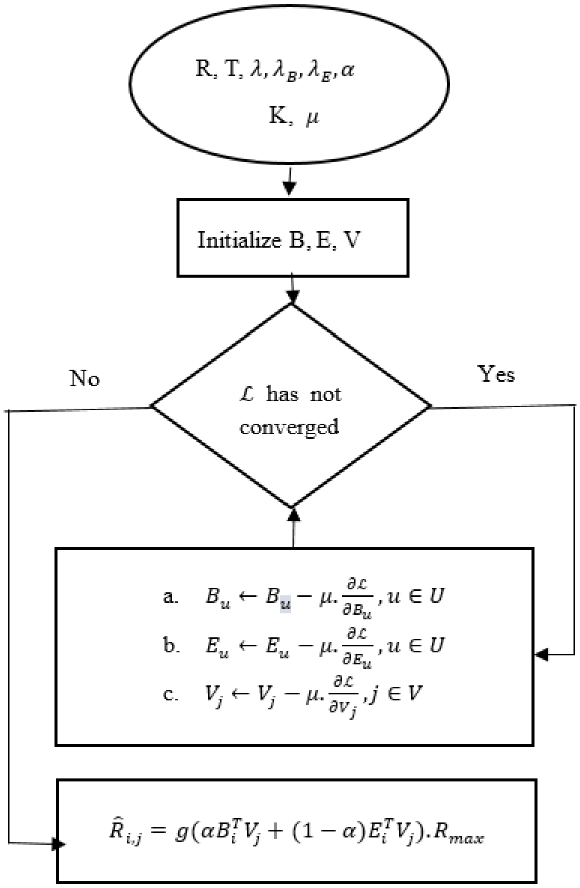

R, T, $\lambda, \lambda_B, \lambda_E, \alpha$
K, μ
Initialize B, E, V
No
Yes
$\mathcal{L}$ has not converged
a. $B_u \leftarrow B_u - \mu . \frac{\partial \mathcal{L}}{\partial B_u}, u \in U$
b. $E_u \leftarrow E_u - \mu . \frac{\partial \mathcal{L}}{\partial E_u}, u \in U$
c. $V_j \leftarrow V_j - \mu . \frac{\partial \mathcal{L}}{\partial v_j}, j \in V$
$\hat{R}_{i,j} = g(\alpha B_i^T V_j + (1-\alpha) E_i^T V_j) . R_{max}$

Fig. 3. A Learning Algorithm for the Proposed Model.

IV. Experimental Evaluation

A. *Description of the Dataset*

Epinions is considered a real-world dataset, publicly available, and widely used to evaluate recommender systems in the literature. Epinions contains users' ratings on items and explicit trust/distrust relationships between users, which is necessary to validate our model. The "Epinions" database used in our experiments was compiled by Paolo Massa in a five week "crawling" process (October / December 2003) for Epinions.com [23] , Table I presents statistics of this database.

TABLE I. DATABASE USED IN EXPERIMENTS

Feature	**Epinions**
Items	139 738
Users	40 163
Ratings	664 824
Trusts	487 183
Trusters	33 960
Trustees	49 288
Intervalle	[1-5]

We have also noticed through a set of experiments that have been carried out that the distributions of trustors and trustees of the "Epinions" database correspond to the "Power-Law" distribution, as is the case. In many social networks, and this is illustrated in the Fig. 4, where only a few trustees have many trustors, while most trustees have only a few trustors. This indicates that there is a significant dislocation in the confidence data provided by users.

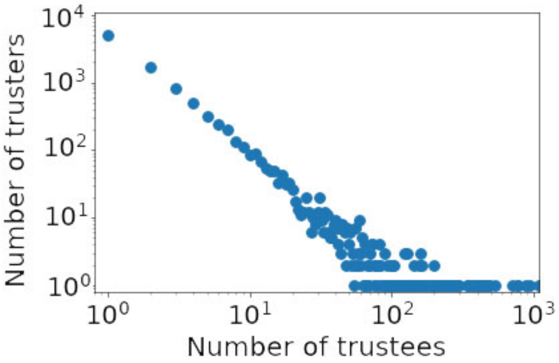


Fig. 4. Distribution of Trustees Relative to Trusters.

B. *Ratings Measures*

To measure the predictive quality of the proposed method compared to collaborative filtering and other methods based on trust, we use two measures [7, 20, 24]:Mean Absolute Error (MAE): it measures the mean absolute differences between the real ratings $R_{i,j}$ given by the user and the ratings predicted by the recommendation system R and is illustrated by the following relation:

$$MAE = \frac{1}{N}\sum_{(u_i,v_j)\in TestSet}\left|\hat{R}_{i,j} - R_{i,j}\right| \quad (13)$$

Where N is the number of ratings we want to test, the lower the MAE measurement value is, the higher the prediction becomes.

RMSE (Root Mean Absolute Error) measures the square root of the mean square of the differences between the actual ratings $R_{i,j}$ given by the user and the ratings predicted by the recommendation system, and is explained by the following relation:

$$RMSE = \sqrt{\frac{1}{N}\sum_{(u_i,v_j)\in TestSet}(\hat{R}_{i,j} - R_{i,j})^2} \quad (14)$$

The lower the RMSE measurement value is, the higher the prediction becomes.

In addition to the ratings measures, it is also necessary to choose the technology in which the recommendation system will be rated. In this paper, we use the technique of "Five-Fold cross-validation" for training and testing, in which the database is divided into two parts. The former part concerns model training, while the latter is concerned with testing for its rating. Specifically, the database is randomly divided into five parts, one part is kept as test data, and the remaining four parts are used as training data. This process is repeated five times. Each of these five parts is used as test data only once, and at the end, the results are averaged to obtain the exact result. In our experiments, in each part, we use 80% of the data as training data and the remaining 20% as the test data, which means that we randomize 80% of the ratings in the "Epinions" database as training data to predict the remaining 20% of ratings.

C. *Experimental Parameters*

To show the performance improvements of the proposed method, we compare it with the traditional matrix analysis method and with a few confidence-based methods found in previous studies that are relevant to our present study.

TABLE II. THE PERFORMANCE OF PROPOSED METHOD

Methods	**Rationale**	**Optimal Parameters**
PMF[25]	The basic matrix analysis method uses the rating matrix only for recommendations, without considering social relationships between users.	$\lambda_u = 0.001$ $\lambda_v = 0.001$
SoRec[8]	The method analyses the rating matrix and the users' social trust matrix by sharing the same space of latent features.	$\lambda_c = 1$ $\lambda_u = 0.001$ $\lambda_v = 0.001$ $\lambda_z = 0.001$
RSTE[11]	The method models users' ratings as a balance between their own preferences and those of their trustee's people.	$\alpha = 0.4$ $\lambda_u = 0.001$ $\lambda_v = 0.001$
SocialMF[12]	The method in which the diffusion of trust is considered when modelling trust relationships between users.	$\lambda_t = 1$ $\lambda_u = 0.001$ $\lambda_v = 0.001$
SoReg[13]	The method in which the trust relations are modelled by supposing that the distance between the latent features of the trustors must be minimal.	$\beta = 0.001$ $\lambda_1 = 0.001$ $\lambda_2 = 0.001$
TrustMF[24]	The method in which users are questioned at two spaces of latent features, the space of the user as trustor and the space of the user as trustee, by analyzing the trust network.	$\lambda_t = 1$ $\lambda = 0.001$
GlotMF	Our proposed method in which the general and the local social context of trust relationships between users is exploited, where local information represents the preferences of the two different states of users (trustee and trustor) which are learned by modelling the explicit and implicit interactions between users, while the general information represents the reputation of the user in the social network.	$\alpha = 0.6$ $\lambda_b = 0.1$ $\lambda_e = 0.1$ $\lambda = 0.001$

To check all the methods that are compared with the methods we use, we define our own optimization parameters according to the corresponding references that are based on our experiments, as shown in the Table II. To compare these methods fairly, we define the dimension K of the latent feature space, for example, five and ten. In reference to all the methods, we adopt the same initialization strategy, which randomly initializes all the latent feature matrices with a distribution uniform in the domain [0, 1].

In this test, we will focus on two user's points of view to measure the performance of the different methods compared:

- Perspective of All Users: users who have at least one rating.
- Perspective of Cold-Start Users: Users with less than 5 ratings.

D. *Experimental Results*

the results of our experiments that we carried out on the "Epinions" database for the "All Users" and "Cold-Start Users" perspectives on the "MAE" and "RMSE" measurements used in the process of evaluation and with different parameters for the dimension of the latent factor space (k = 5, k = 10). The experimental results of our proposed method and the methods we compared with, are presented in the Table III and Table IV.

TABLE III. RECOMMENDATION PERFORMANCE COMPARISONS ON EPINIONS DATA SET, THE CASE OF "ALL USERS"

Metrics	**MAE**		**RMSE**	
	K=5	K=10	K=5	K=10
PMF	0.979	0.909	1.290	1.197
RSTE	0.950	0.958	1.196	1.278
SoRec	0.882	0.884	1.114	1.142
SoReg	0.994	0.932	1.315	1.232
SocialMF	0.825	0.826	1.070	1.082
TrustMF	0.818	0.820	1.069	1.095
GlotMF	0.804	0.805	1.043	1.044

TABLE IV. RECOMMENDATION PERFORMANCE COMPARISONS ON EPINIONS DATA SET, THE CASE OF "COLD-START USERS"

Metrics	**MAE**		**RMSE**	
	K=5	K=10	K=5	K=10
PMF	1.451	1.153	1.770	1.432
RSTE	1.051	0.981	1.266	1.313
SoRec	0.892	0.846	1.138	1.180
SoReg	1.398	1.139	1.735	1.437
SocialMF	0.884	0.857	1.133	1.152
TrustMF	0.891	0.853	1.125	1.176
GlotMF	0.868	0.868	1.105	1.108

1) *Analysis and discussion of the results*

- The performance of the traditional "PMF" method is lower than the performance of other compared methods based on a matrix analysis which, in our opinion, benefits from the relationships of trust in the two perspectives "All Users" and "Cold-Start Users" which underline the importance of trust in improving the performance of model-based recommendation systems.
- From an "All Users" perspective, the TrustMF and GlotMF models perform better than other trust-based models that map users to a single latent feature space. In addition, our proposed model "GlotMF" outperforms the "TrustMF" model among all models compared by 1.4% and 2.6% respectively when K = 5, and by 1.5%, 5.1% when K = 10 in terms of "MAE" and "RMSE" respectively.
- From a Cold-Start User perspective, there is no single model that works best among trust-based models. In general, our suggested model performs better than the others. Although, with a few noted exceptions in terms of MAE, our model is more robust in terms of RMSE. Since all confidence-based models aim to improve the squared errors between predicted values and actual values. Whereas, the "RMSE" measurement is more significant than the "MAE" measurement (where "RMSE" calculates the square of the differences between the actual ratings and the predicted ratings, thusly, penalizing large errors more than the "MAE"), and, therefore, our model always has the best performances, surpassing the "TrustMF" model the best among the compared models at the level of the "RMSE" measure when K = 5, and the "SocialMF" model the best at the level of the "RMSE" measure when K = 10.
- In addition, we notice that the "GlotMF" model achieves better performance in both perspectives, which confirms the efficiency of taking the implicit links between the users beside the explicit links in improving the performance of the recommendation, as well as exploiting the general social context of users contributed to our model outperforming other models based on trust.
- Although the percentage of relative improvements in our model compared to other compared models is small, these improvements are significant, as researchers in [4, 26] noted that even small improvements in MAE and RMSE measurements can lead to significant differences in the recommendations. As evidence to the above-mentioned result, a million-dollar award was submitted to the Netflix Prize competition in October 2006 for a 10% improvement in the RMSE metric over traditional sponsorship methods.

2) Checking the performance of the proposed model on users with different degrees of confidence: other series of experiments to verify the performance of the proposed model for users with different degrees of trust relationships was carried out to compare our method with other trust-based methods and test the possibilities of these different methods by benefiting trust data for the recommendation. The degrees of trust relationships can be defined as the total of the trustees' neighbors of the user and the trustors' neighbors of the user themselves. To conduct our experiments, we first group all the users into several groups (up to seven groups) according to their degrees of trust, these groups are "1-5", "6-10", "11-20 "," 21-40 "," 41-100 "," 101-500 ", and "> 500 "as used by the researchers. Then, we calculate the predictive error in each group respectively in terms of measures" MAE "and" RMSE "when the number of latent features is equal to five and ten, the results of these experiments are shown in the Fig. 5, Fig. 6, Fig. 7, and Fig. 8.

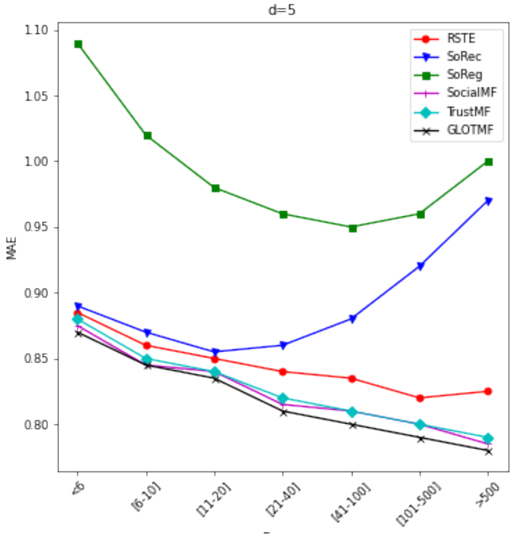


Fig. 5. The Predict Error on users in different Methods when d=**5** (MAE).

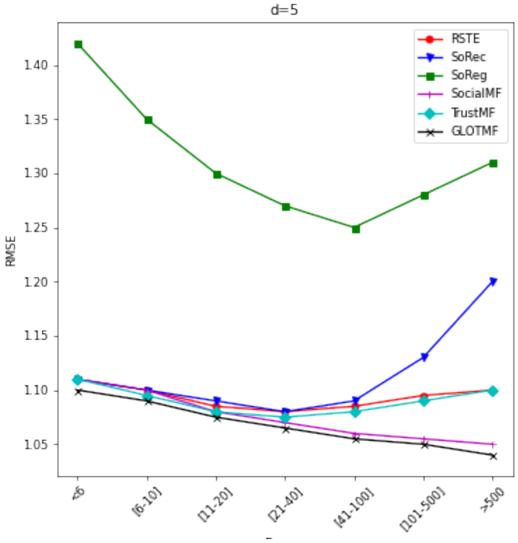


Fig. 6. The Predict Error on users in different Methods when d=**5** (RMSE).

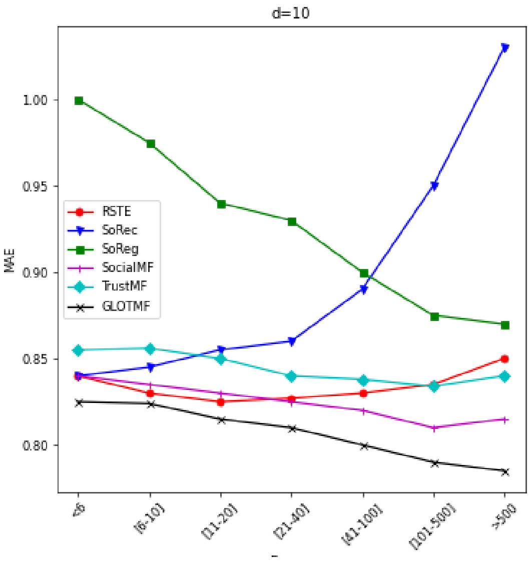


Fig. 7. The Predict Error on users in different Methods when d=10 (MAE).

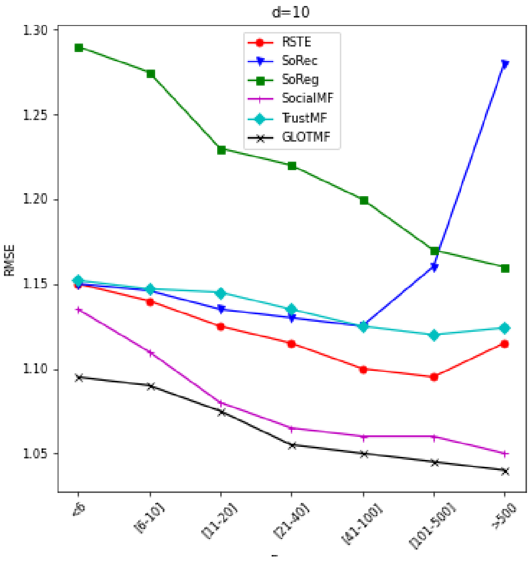


Fig. 8. The Predict Error on users in different Methods when d=10 (RMSE).

The results obtained show that the performance of the six compared methods differ to some extent from the different trust groups, and our "GlotMF" model works stable and shows the best quality for most groups, especially for the group with five relationships of maximum trust (about 63.4% of users), and for the group of 6 to 10 trust relationships (about 11.6% of users) on the MAE and RMSE measures in the two dimensions five and ten. This strongly suggests that our model can benefit from dispersed trust data more effectively than other trust-based collaborative filtering methods.

V. Conclusion

The Research tried to present our proposal to improve the accuracy of recommendations in the model-based collaborative filtering method by taking advantage of trust relationships between users. This is done by presenting a new data fusion strategy rating and trust data to express users' mutual impact on their opinions more logically, and, thusly, predict the unknown values of user ratings on items more accurately and efficiently. In our proposed method, local information represents the preferences of the two different user states trustee and trustor which are learned by modelling the explicit and implicit interactions between users. While general information represents the reputation of the user in the social network. Experiments were performed on a real database, "Epinions", and in two user perspectives, "All Users" and "Cold-Start Users". The results of these experiments showed that the suggested method brought up significant improvements in terms of predictive accuracy, and its effectiveness in alleviating data dispersion and cold start problems compared to other methods based on models that take advantage of trust relationships to improve the accuracy of recommendations. Current work relies on the explicit trust granted explicitly by users, but the user might refute to share or disclose this information, for example, due to privacy concerns. Moreover, many of these available datasets contain explicit trust information. Therefore, implicit trust values can be inferred from user behaviors to improve the generalizability of proposed method. In future work, we intend to improve the proposed model, and it will be interesting to study more extensions.

References

[1] C. Gao, W. Lei, X. He, M. de Rijke, and T.-S. Chua, "Advances and challenges in conversational recommender systems: A survey," arXiv preprint arXiv:2101.09459, 2021.

[2] L. Huang, H. Ma, X. He, and L. Chang, "Leveraging Multisource Information in Matrix Factorization for Social Collaborative Filtering," in 2020 International Joint Conference on Neural Networks (IJCNN), 2020: IEEE, pp. 1-8.

[3] S. Natarajan, S. Vairavasundaram, S. Natarajan, and A. H. Gandomi, "Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data," Expert Systems with Applications, vol. 149, p. 113248, 2020.

[4] D. P. D. Rajendran and R. P. Sundarraj, "Using topic models with browsing history in hybrid collaborative filtering recommender system: Experiments with user ratings," International Journal of Information Management Data Insights, vol. 1, no. 2, p. 100027, 2021.

[5] P. Ma, L. Wang, and J. Qin, "A Low-Rank Tensor Factorization Using Implicit Similarity in Trust Relationships," Symmetry, vol. 12, no. 3, p. 439, 2020.

[6] C. Park, D. Kim, J. Oh, and H. Yu, "Improving top-K recommendation with truster and trustee relationship in user trust network," Information Sciences, vol. 374, pp. 100-114, 2016.

[7] B. Yang, Y. Lei, J. Liu, and W. Li, "Social collaborative filtering by trust," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 8, pp. 1633-1647, 2016.

[8] F.-S. Hsieh, "Trust-based recommendation for shared mobility systems based on a discrete self-adaptive neighborhood search differential evolution algorithm," Electronics, vol. 11, no. 5, p. 776, 2022.

[9] Y. Ruan and A. Durresi, "A survey of trust management systems for online social communities–trust modeling, trust inference and attacks," Knowledge-Based Systems, vol. 106, pp. 150-163, 2016.

[10] M. Al-Ghamdi, H. Elazhary, and A. Mojahed, "Evaluation of Collaborative Filtering for Recommender Systems," International Journal of Advanced Computer Science and Applications, vol. 12, no. 3, 2021.

[11] C. Xu, A. S. Ding, and K. Zhao, "A novel POI recommendation method based on trust relationship and spatial–temporal factors," Electronic Commerce Research and Applications, vol. 48, p. 101060, 2021.

[12] J. Shokeen, C. Rana, and P. Rani, "A trust-based approach to extract social relationships for recommendation," in Data Analytics and Management: Springer, 2021, pp. 51-58.

[13] K. Zhang, X. Liu, W. Wang, and J. Li, "Multi-criteria recommender system based on social relationships and criteria preferences," Expert Systems with Applications, vol. 176, p. 114868, 2021.

[14] L. Yang and X. Niu, "A genre trust model for defending shilling attacks in recommender systems," Complex & Intelligent Systems, pp. 1-14, 2021.

[15] J. Shokeen and C. Rana, "Social recommender systems: techniques, domains, metrics, datasets and future scope," Journal of Intelligent Information Systems, vol. 54, no. 3, pp. 633-667, 2020.

[16] H. Zhang, I. Ganchev, N. S. Nikolov, and M. Stevenson, "UserReg: A simple but strong model for rating prediction," in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021: IEEE, pp. 3595-3599.

[17] A. Rahim, M. Y. Durrani, S. Gillani, Z. Ali, N. U. Hasan, and M. Kim, "An efficient recommender system algorithm using trust data," The Journal of Supercomputing, vol. 78, no. 3, pp. 3184-3204, 2022.

[18] W. Zhang, X. Zhang, H. Wang, and D. Chen, "A deep variational matrix factorization method for recommendation on large scale sparse dataset," Neurocomputing, vol. 334, pp. 206-218, 2019.

[19] C. N. Mabude, I. O. Awoyelu, B. O. Akinyemi, and G. A. Aderounmu, "An Integrated Approach to Research Paper and Expertise Recommendation in Academic Research," International Journal of Advanced Computer Science and Applications, vol. 13, no. 4, 2022.

[20] Y. Pan, F. He, H. Yu, and H. Li, "Learning adaptive trust strength with user roles of truster and trustee for trust-aware recommender systems," Applied Intelligence, vol. 50, no. 2, pp. 314-327, 2020.

[21] R. Chen, Y.-S. Chang, Q. Hua, Q. Gao, X. Ji, and B. Wang, "An enhanced social matrix factorization model for recommendation based on social networks using social interaction factors," Multimedia Tools and Applications, pp. 1-31, 2020.

[22] S. Xu, H. Zhuang, F. Sun, S. Wang, T. Wu, and J. Dong, "Recommendation algorithm of probabilistic matrix factorization based on directed trust," Computers & Electrical Engineering, vol. 93, p. 107206, 2021.

[23] S. Liu, L. Zhang, and Z. Yan, "Predict pairwise trust based on machine learning in online social networks: A survey," IEEE Access, vol. 6, pp. 51297-51318, 2018.

[24] B. Yang, Y. Lei, J. Liu, and W. Li, "Social Collaborative Filtering by Trust," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 8, pp. 1633-1647, 2017, doi: 10.1109/TPAMI.2016.2605085.

[25] R. Gund, J. Andro-Vasko, D. Bein, and W. Bein, "Recommendation System Using MixPMF," in ITNG 2022 19th International Conference on Information Technology-New Generations, 2022: Springer, pp. 263-268.

[26] T. Anwar and V. Uma, "Comparative study of recommender system approaches and movie recommendation using collaborative filtering," International Journal of System Assurance Engineering and Management, vol. 12, no. 3, pp. 426-436, 2021.

Impervious Surface Prediction in Marrakech City using Artificial Neural Network

Sulaiman Mahyoub[1], Hassan Rhinane[2], Mehdi Mansour[3], Abdelhamid Fadil[4], Waban Al okaishi[5]

Laboratory Geosciences (of Department of Geology) Faculty of Sciences, Hassan II University, Casablanca, Morocco[1, 2, 3]
Hassania School of Public Works, Casablanca, Morocco[4]
Laboratory LTI Lab, Faculty of Sciences Ben M'sik, Hassan II University, Casablanca, Morocco[5]

***Abstract*—Determining an impervious surface is one of the most important topics of remote sensing because of its great role in providing information that benefits decision-makers in urban planning, sustainable development goals, and environmental protection. In recent years, a great development in this field has occurred due to the huge improvement in the algorithms and techniques that are used to map impervious surfaces. In this paper, the deep learning technique has been implemented to investigate the extraction of impervious surfaces in Marrakesh city, based on Landsat images. 9000 polygons and 13840 points have been taken to prepare label data by random forest in Google Earth Engine (GEE). In addition, all preprocessing steps for remote sensing images have been implemented in GEE. An artificial neural network (ANN) has been used to determine impervious surfaces. After training and testing the proposed network on Landsat image datasets, precision, accuracy, recall, and F1-score matrix scores were 0.79, 0.98, 0.87, and 0.82, respectively. The experimental results show that this method is efficient and precise for mapping the impervious surfaces of Marrakesh city.**

***Keywords*—Deep-learning; remote sensing; artificial neural network ANN; impervious surface**

I. Introduction

Impervious surfaces are a prominent indicator of human presence and are characterized as land covers that have been contained by manmade structures that restrict water penetration into the soil, such as building roofs, roadways, and parking lots [1]. The growth of impervious surfaces and urban areas causes a variety of problems, including environmental risks and sociopolitical consequences [2, 3]. In order to achieve sustainable development goals, urban planning, and environmental protection, it is necessary to provide all details and specific information that cover the long-term dynamics of impervious surfaces [4, 5]. Since the 1970s, remote sensing instruments have been widely employed to monitor the Earth. Everywhere across the world, day or night, these satellites gather data at a low cost or for free. With this, the quantity of data that can be used to research impervious surfaces has increased significantly [6-8]. In order to successfully extract information from such a massive volume of data, high-performance processing skills, such as machine learning and cloud computing platforms, are required. These algorithms and technologies have been shown to be extremely efficient [9-12].

Remote sensing has been widely utilized to identify impervious surfaces. Several articles have been published describing the state-of-the-art of this topic [13-21]. Earlier, statistical remote sensing indices including the Normalized Difference Built-up Index (NDBI) [22], Normalized Difference Impervious Surface Index (NDISI) [23], modified NDISI [24], and perpendicular impervious surface index(PISI) [25] , biophysical composition index (BCI) [26] , and the normalized difference vegetation index (NDVI) have been developed to map impervious surfaces. Scholars and researchers have moved their focus to machine learning techniques such as Random Forest [27], Support Vector Machine (SVM) [28], and Classification and regression trees (Cart) [29], among others. Similar to our earlier study, this is a continuation of that work [30]. In recent years, deep learning algorithms and cloud computing infrastructures have witnessed tremendous development, resulting in significant advancements in image processing applications such as classification, segmentation, and change detection [31-33]. As a result, the remote sensing community is attempting to modernize remote sensing image preprocessing and processing in order to keep up with this development.

One of these major accomplishments is the Google Earth Engine cloud computing platform, which is free and combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities, as well as its community [34, 35]. This is constantly updated and has many scientists, researchers, and developers working on it to develop remote sensing and environmental applications. Despite the fact that deep learning has unique automatic feature learning capabilities and strong nonlinear complex function expression and fitting capabilities, and it can combine low-level features to form more abstract high-level representations and attribute categories or features, it requires significantly more training data than conventional machine learning supervised classifications [36].

In this work, deep learning techniques have been proposed to determine impervious surfaces of Marrakesh city, based on Landsat images to take advantage of the large archive owned by this satellite (1972–now) [37, 38]. 6787 polygons have been extracted from the cadastral plan, and 13122 labeled points have been chosen carefully from high resolution images to prepare label data images by random forest in the Google Earth engine (GEE). In addition, all preprocessing for remote sensing images for training and testing images has been implemented in GEE. An artificial neural network (ANN) has been used to implement this task. The remainder of this paper is organized as follows: Section II discusses the pieces that are related to this topic. The proposed methodology for predicting impervious surfaces for the city of Marrakech has been illustrated in Section III. Section IV describes the experimental results and their evaluations. The conclusion is found in Section V.

II. Related Work

In general, the literature on remote sensing techniques for detecting impervious surfaces is divided into three categories: statistical-index based, machine learning based.

1) Statistical-index based: In this subsection, index-methods will be discussed briefly. These statistical indicators have been modeled by calculating two or more bands in order to improve the spectrum of specific features of the target areas, such as NDBI, modified NDISI, NDIS, PISI, and BCI. The computing formulas for some of these indicators are presented below:

$$NDBI = \frac{SWIR1 - NIR}{SWIR1 + NIR}$$

$$\mathrm{MNDWI} = \frac{\mathrm{G - SWIR1}}{G + SWIR1}$$

$$\mathrm{NDISI} = \frac{\mathrm{Tb + (MNDWI + NIR + SWIR1)/3}}{\mathrm{Tb - (MNDWI + NIR + SWIR1)/3}}$$

$$\mathrm{PISI} = 0.8192 * \mathrm{B} - 0.5735 * \mathrm{NIR} + 0.075$$

$$NDVI = \frac{NIR\ - RED}{NIR\ + RED}$$

where:

- SWIRI is the band of shortwave infrared waves (Landsat 8 band 6).while.
- Tb indicated to the temperature of the TIRS1 thermal band's brightness.
- NIR refers to the near-infrared band pixel values (Landsat 8 band 5).
- RED, G, and B represent the red, green and blue canals on Landsat 8, also known as Bands 4, 3 and 2 respectively.

Despite these statistical-based algorithms being obvious and simple to use, they faced a few restrictions in their performance. For example, in regions with large concentrations of topsoil, NDBI does not work well because it cannot distinguish between urbanized areas and vegetation. On the other hand, the NDISI relied on land-surface temperature, which might fail in zones where the heat island effect is less prominent. Furthermore, the Tb band is also problematic since it is not present in all sensors. In general, it's hard to use these indicators alone to predict the impervious surface because the threshold interval changes from zone to zone and hasn't been adjusted yet [25].

a) Texture-based methods: Previously, scholars and researchers have extracted the impervious surface using various methods based on spectral analysis and mixed pixel decomposition to examine the variations in reflectance features of impervious surface areas using medium-to-low-resolution remote sensing images, such as Landsat [39, 40].

b) Machine learning based: Machine learning algorithms have been extensively conducted to model impervious surfaces, such as the maximum likelihood classifier, support vector machines (SVM), classification and regression trees (CART) algorithm, and random forest. In spite of the fast development of these approaches, they need vast quantities of labeled training and assessment data. Nonetheless, the gathering and generation of this quantity of labeling data by visual interpretation is regarded as time-consuming and cost-intensive [28]. In addition, the machine learning approach has been challenged by the spectral and textural complexity of impervious surface area (ISA) in mixed pixels, including different combinations of ISA and other land cover types [41].

III. Research Methodology

Marrakesh is one of the most important Moroccan cities, it is located at 31° 37′ 48″ N 8° 0′ 32″ W, and has a surface area of more than 230 km²; It suffers from the effects of migration from the surrounding countryside to it; it was chosen by the atlas project (www.atlasofurbanexpansion.org/) to study urban expansion from 1985 to the present; and it was also chosen in our previous work [30]. This paper will determine the impervious surface of Marrakesh city using deep learning technique.

A. Database Preparation for Learning and Testing the Proposed Model

Cloud computing is very important for deep learning applications since it provides unlimited storage, provisioning, and updating, as well as guaranteed privacy and security [42, 43]. Users of cloud services can also improve server usage, dynamic scalability, and reduce the time for creating new applications [44]. The Google Earth engine platform (GEE) is utilized to implement all steps for training and testing the deep learning model.

In this study, our goal is to use deep learning techniques to extract the impervious surface of the city of Marrakech based on Landsat images. 6787 polygons and 13122 points have been chosen to prepare labeling data by using random forest in GEE. These points had been collected using visual interpretation analysis, which is based on our power to match patterns and colors in an image to real-world features, from high resolution images from different resources such as (Google earth, Sentinel-2 OSM). According to the polygons, they were taken from plan cadastral of Marrakesh city. Then these data are uploaded to GEE cloud platform as a form of features collections. After that, a sampling operation was performed to link the features (Landsat image datasets 8 bands: B2, B3, B4, B5, B6, B7, NDVI, and NDBI of each point and polygon with the corresponding pixels in the target image (Marrakesh city image).

In the present study, random forest algorithm has utilized to produce labeled data image as it is the best obtained result in our previous study [30]. Fig. 1 shows the resulted image from random forest, the white pixels represent the impervious surface of Marrakesh city.

Random forest is composed of many decision trees. Furthermore, it employs randomness to improve accuracy and avoid over fitting, which can be a significant problem for such a complex algorithm. These methods generate decision trees from a random sampling of data and extract predictions from each tree. Following that, they vote on the most possible solution. The use of random forest (RF) for image classification has gained popularity due to its ease of use (e.g. relatively insensitive to classification parameters) and generally high accuracy [45]. This algorithm works as follows: Assuming the dataset has "m" features, the random forest will select "k" features randomly, where k < m. The algorithm will now select the node with the biggest information gain among the k features as the root node. The algorithm then divides the node into child nodes and continues the procedure "n" times more; a forest with n trees is obtained. Finally, bootstrapping, which entails combining the outcomes of all of the decision trees in random forest, will be performed. Because it is based on the functionality of decision trees, it is definitely one of the most complicated algorithms[46].

After obtaining our labeled image, it is necessary to prepare the database (the labeled image) to be the input of the ANN model. The steps for preparing the database are as follows:

- First step is to flattening the data shape from multi-dimensions (Landsat image datasets eight layers: b2, b3, b4, b5, b7, b7, NDVI, NDBI) into two dimensions array. This is a very important step because Flatten makes the serialization of a multidimensional tensor apparent (typically the input one here eight layers). The mapping between the (flattened) input tensor and the first hidden layer is now possible. Each member of the (serialized) input tensor will be linked with each element of the hidden array if the first hidden layer is "dense". The way the input tensor is mapped into the first hidden layer would be unclear if you didn't utilize Flatten. Fig. 2 shows the result of the flatting operation.
- The second step is data cleaning, by removing no-data pixels from the arrays.
- The third is to divide the data by 60 percent for training and 40 percent for validation accordingly.
- The fourth stage is normalization, which is required by many machine learning methods, including NNs. This indicates that the histogram has been stretched and scaled to meet a particular range of values (here, 0 to 1). We will normalize our features to fulfill this requirement. Normalization can be performed by subtracting the minimum value and dividing it by the range. Because the Landsat 8 data is a 16 bits data (216= 65536 values), the lowest and highest values are 0 and 65536. Fig. 2 also shows the result of the normalization operation.
- Reshaping the features from two to three dimensions, such that each row represents an individual pixel, is another pre-processing step. Fig. 3 indicates the result of converting the features from two to three dimensions.

B. Architecture of the Proposed Model

In recent years, an artificial neural network (ANN) has gotten a lot of attention because of its ability to solve complicated nonlinear problems in domains like computer vision, image processing, natural language processing, and so on. For this reason, an ANN deep learning model has been implemented to improve impervious surface extraction for Marrakech, Morocco.

The basic architecture of an ANN model consists of input and output layers as well as hidden layers with various weights. Then, these several layers are connected by neurons. Every single neuron is inspired by a biological neuron, and produces a single output, and is known as a perceptron.

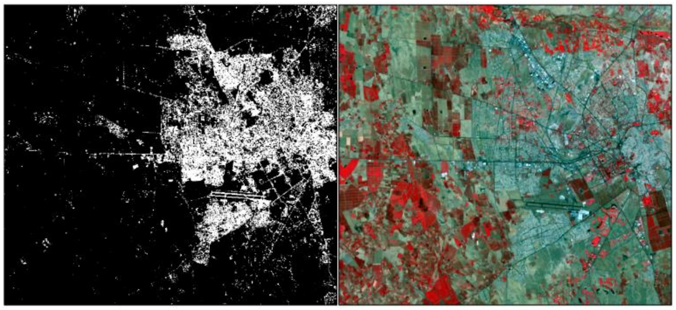


Fig. 1. Labeled Image and the Original Image.

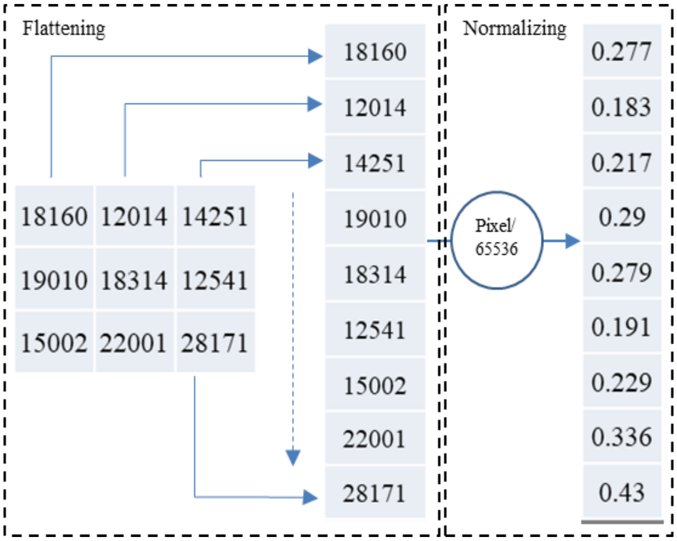
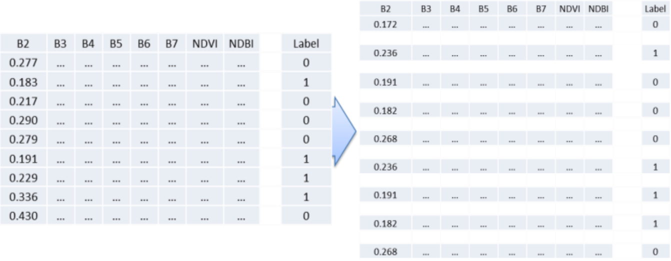


Fig. 2. Flattening and Normalizing.



B2	B3	B4	B5	B6	B7	NDVI	NDBI		Label
0.277	...	...	...	...	...	...	...		0
0.183	...	...	...	...	...	...	...		1
0.217	...	...	...	...	...	...	...		0
0.290	...	...	...	...	...	...	...		0
0.279	...	...	...	...	...	...	...		0
0.191	...	...	...	...	...	...	...		1
0.229	...	...	...	...	...	...	...		1
0.336	...	...	...	...	...	...	...		1
0.430	...	...	...	...	...	...	...		0

B2	B3	B4	B5	B6	B7	NDVI	NDBI		Label
0.172	...	...	...	...	...	...	...		0
0.236	...	...	...	...	...	...	...		1
0.191	...	...	...	...	...	...	...		0
0.182	...	...	...	...	...	...	...		0
0.268	...	...	...	...	...	...	...		0
0.236	...	...	...	...	...	...	...		1
0.191	...	...	...	...	...	...	...		1
0.182	...	...	...	...	...	...	...		1
0.268	...	...	...	...	...	...	...		0

Fig. 3. Reshaping.

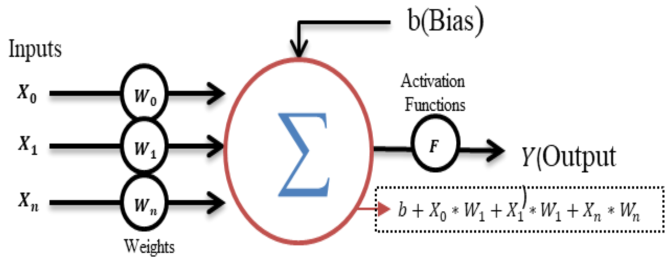


Fig. 4. Basic Architecture of ANN.

The perceptron's basic architecture is illustrated in the diagram above (Fig. 4). Initially, X0, X1,... and Xn indicate inputs. The inputs (X0, X1,.. and Xn) are multiplied by a connection of weights (W0, W1,.. and Wn) and then summed with the bias value b (which permits the activation function to be shifted up or down). To obtain the output of the perceptron, the output of the summation operation is applied to the activation function. If no activation function is used, the output signal is just a linear function, and the model will be unable to learn and model complex data (such as images, videos, audio, speech, and so on). A neural network needs to learn and represent anything, as well as any arbitrarily complicated function that links an input to an output. Therefore, a neural network has to use a non-linear activation function to perform this task.

C. Implementation of ANN

In our model, the input variables are Landsat 8 image bands from Band 2 – Blue to Band 7 – SWIR 2, NDVI, and NDBI. We chose the additional bands (NDVI and NDBI) to increase the amount of data, which leads to improving the performance of our model.

The proposed model has two hidden layers; the first layer contains 32 neurons, while the other contains 16 neurons, followed by an output layer with two classes, impervious and non-impervious pixels, as shown in Fig. 4. The Softmax activation function is used because it is the most suited function for our situation to estimate impervious surface pixels.

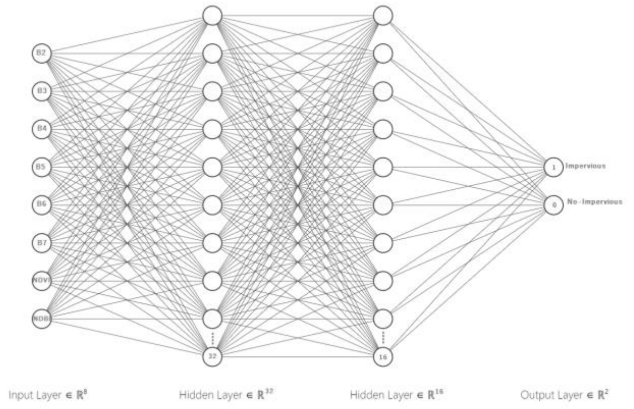

B2
B3
B4
B5
B6
B7
NDVI
NDBI
32
16
1
Impervious
0
No -Impervious
Input Layer ∈ $\mathbb{R}^8$
Hidden Layer ∈ $\mathbb{R}^{32}$
Hidden Layer ∈ $\mathbb{R}^{16}$
Output Layer ∈ $\mathbb{R}^2$

Fig. 5. Model Architecture.

IV. Results and Discussion

The impervious surface classification results will be discussed in this section. To detect the impervious surfaces, a series of preprocessing steps have been implemented, as already mentioned in the previous section. We utilized an ANN model to implement the detection of impervious surfaces. This model was trained and tested using a dataset, which is taken from the Landsat-8 archive. This dataset has been corrected, filtered, and exported in the form of a Geographic Tagged Image File Format (GeoTIFF) Image (Fig. 1) in GEE. This image contains eight bands with 30 meter spatial resolution, and its dimensions are 961 pixels in height and 2476 pixels in width. To train and test the proposed model, the database is split into two parts; the first contains 60% of the database, which is used to train the model, while the other part is used for testing the model.

A. Performance Accuracy Assessment of ANN Model

At this stage, the proposed model has been implemented on the Google Colaboratory Cloud platform. This platform facilitates the training of deep learning models online by providing a lot of computing power, better than our local machines. After training and testing our model, we obtained the confusion matrix's components as shown in Table I. The explanation of the confusion matrix is as follows:

- True Positive means that the pixel is predicted as an impervious surface pixel and it's true.
- True Negative represents the pixels that are not impervious surface pixels, and it's true.
- False Positive means that the pixel is predicted as an impervious surface pixel, but it’s false.
- False Negative means that the pixel is predicted as a non-impervious surface pixel, but it’s false.

Table I indicated that, our model has very high performance as the positive impervious surface detection is very high and the wrong detection is low. To quantitatively evaluate our model performance, four measurements are utilized, namely accuracy, recall, precision, and F1-score.

Accuracy: The discernment required to distinguish between impervious and non-impervious surfaces. The percentage of all assessed instances that are true positive and true negative is derived to assess a test's accuracy. It can be expressed mathematically as:

$$\text{Accuracy} = \frac{(TN + TP)}{(TN + TP + FN + FP)}$$

TABLE I. THE OBTAINED RESULTS OF OUR MODEL

confusion matrix		
Output -predictable image	***Actually Positive(1)***	***Actually Negative(0)***
Impervious surface Class - 1	True positive (TP) **916106**	False positive (FP) 6778
Non-Impervious surface class -0	False negative (FN) 3873	True negative (TN) 25018

That is the total number of assessments divided by the number of correct assessments. Additionally, precision, recall, and F1 score are used to assess the proposed methodology of performance. Precision is defined as the proportion of accurately predicted positive observations to all positive observations. A low proportion of false positives is associated with high precision. It can be expressed as follows:

$$\text{Precision } = \frac{\text{TP}}{\text{TP} + \text{FP})}$$

Recall: The recall is the proportion of correctly foreseen positive observations to all of the actual class observations. To put it another way, it can be expressed as.

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

F1 Score: This is the recall and precision weighted average. As a result, both false positive and false negative numbers are taken into account. In situations where the distribution class is unbalanced, F1 is more helpful than accuracy. The formula for this is F1.

$$\text{F1 Score} = \frac{2(\text{Recall} * \text{Precision})}{(\text{Recall} * \text{Precision })}$$

Table II shows the obtained results of the four measurement parameters that have been chosen.

TABLE II. THE OBTAINED RESULTS OF THE FOUR MEASUREMENTS STATISTICAL

Accuracy	*Precision*	*Recall*	*F1 Score*
0.98%	0.79%	0.87%	0.82%

It can be noted from Table II that our method has high performance in all measurement parameters (accuracy, precision, recall, and F1-score) as their values are higher than 0.79%. This is because of the good design of our model architecture and the large amount of data that is used for training and testing our model.

The evaluation of our ANN model has been applied to a dataset subset covering 124 km on the western side of Marrakesh city, and its dimensions are 589 pixels in height and 1104 pixels in width. Fig. 5 shows the original image of the target area. Fig. 6 presents the output predicted image of our ANN model. Fig. 7 presents the high-resolution image from Sentinel-2 for the same area as the impervious surfaces are clearly monitored. It is noted in Fig. 6 that the impervious surfaces are represented by white pixels. In comparison with Fig. 7, we notice that the impervious surfaces are classified and detected well. Our ANN model has a high accuracy in detecting impervious surfaces.

To assess and analyze the benefits and drawbacks of the suggested model and obtain a comprehensive overview of our research region, high-resolution images from Sentinel-2 have been brought, with true color composited. Fig. 8 displays this image, which describes urbanized Lands located at the end and fringes of the city, the western side of the image, and some settlements in the south of the image.

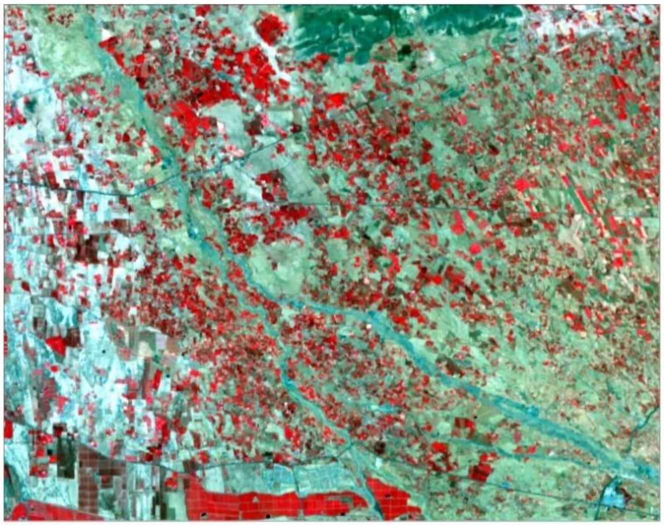


Fig. 6. Marrakech Original Image with Near-Infrared (NIR) Composite Landsat.

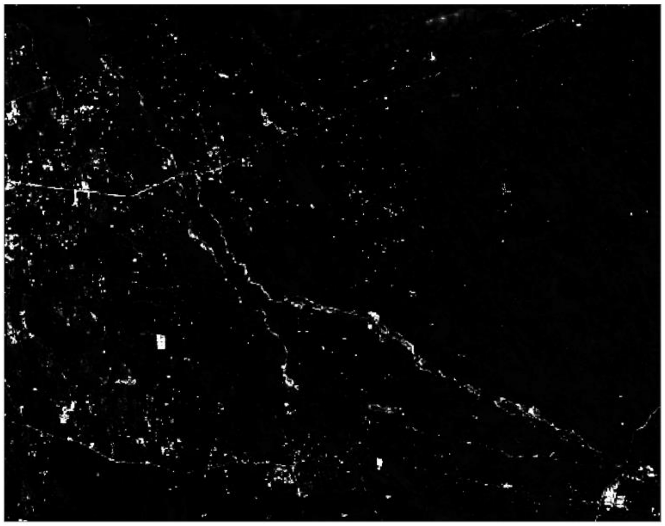


Fig. 7. Output - Predicted Image.

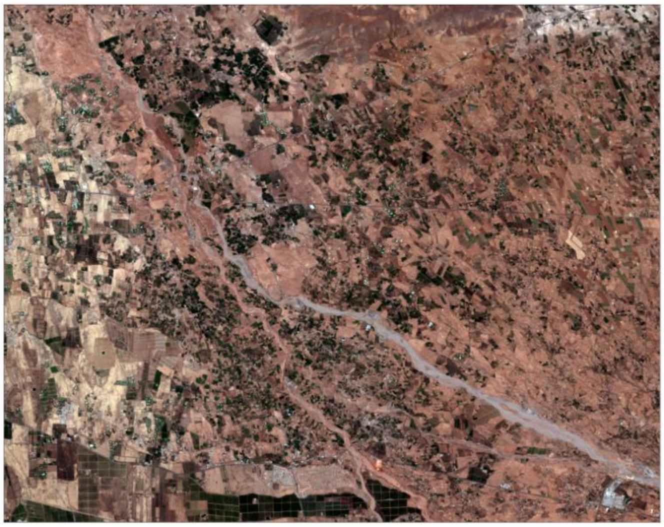


Fig. 8. Subset from Sentinel-2 Image for the Same Area.

V. Conclusion

In this paper, an artificial neural network deep learning model has been used to predict impervious surfaces in Marrakech city; the model is trained and tested using Landsat images. The experimental results show that our model has an accuracy of 81.80% in the precision matrix and 83% in the recall. This is because of the good design of our model architecture and the large amount of data that is used for training and testing our model. In the future, our goal might be to extend the proposed model to a large database containing many cities over many years. Therefore, increasing the amount of data will improve the accuracy of the model.

References

[1] E. B. De Colstoun et al., "Documentation for the global man-made impervious surface (GMIS) dataset from landsat," 2017.

[2] P. Kang, W. Chen, Y. Hou, and Y. J. S. r. Li, "Spatial-temporal risk assessment of urbanization impacts on ecosystem services based on pressure-status-response framework," vol. 9, no. 1, pp. 1-11, 2019.

[3] R. Xu, J. Liu, and J. J. S. Xu, "Extraction of high-precision urban impervious surfaces from sentinel-2 multispectral imagery via modified linear spectral mixture analysis," vol. 18, no. 9, p. 2873, 2018.

[4] M. I. H. Reza and S. A. J. E. i. Abdullah, "Regional Index of Ecological Integrity: A need for sustainable management of natural resources," vol. 11, no. 2, pp. 220-229, 2011.

[5] C. Liu et al., "Arctic's man-made impervious surfaces expanded by over two-thirds in the 21st century," 2022.

[6] C. Liu et al., "An efficient approach to capture continuous impervious surface dynamics using spatial-temporal rules and dense Landsat time series stacks," vol. 229, pp. 114-132, 2019.

[7] M. Feng and X. J. S. B. Li, "Land cover mapping toward finer scales," vol. 65, no. 19, pp. 1604-1606, 2020.

[8] W. Kuang et al., "Global observation of urban expansion and land-cover dynamics using satellite big-data," 2021.

[9] K. N. Markert et al., "Comparing Sentinel-1 Surface Water Mapping Algorithms and Radiometric Terrain Correction Processing in Southeast Asia Utilizing Google Earth Engine," vol. 12, no. 15, p. 2469, 2020.

[10] A. Poortinga et al., "Predictive Analytics for Identifying Land Cover Change Hotspots in the Mekong Region," vol. 12, no. 9, p. 1472, 2020.

[11] K. Phongsapan et al., "Operational Flood Risk Index Mapping for Disaster Risk Reduction Using Earth Observations and Cloud Computing Technologies: A Case Study on Myanmar," (in English), Original Research vol. 7, 2019-December-11 2019.

[12] A. Poortinga et al., "An Operational Before-After-Control-Impact (BACI) Designed Platform for Vegetation Monitoring at Planetary Scale," vol. 10, no. 5, p. 760, 2018.

[13] E. T. Slonecker, D. B. Jennings, and D. Garofalo, "Remote sensing of impervious surfaces: A review," Remote Sensing Reviews, vol. 20, no. 3, pp. 227-255, 2001/08/01 2001.

[14] N. Khanal et al., "A Comparison of Three Temporal Smoothing Algorithms to Improve Land Cover Classification: A Case Study from NEPAL," vol. 12, no. 18, p. 2888, 2020.

[15] Y. Wang and M. Li, "Urban Impervious Surface Detection From Remote Sensing Images: A review of the methods and challenges," IEEE Geoscience and Remote Sensing Magazine, vol. 7, no. 3, pp. 64-93, 2019.

[16] Z. Liu, Y. Wang, and J. J. P. i. G. Peng, "Remote sensing of impervious surface and its applications: a review," vol. 29, no. 9, pp. 1143-1152, 2010.

[17] H. Zhang, H. Lin, Y. Zhang, and Q. Weng, Remote sensing of impervious surfaces in tropical and subtropical areas. CRC Press, 2015.

[18] M. E. Bauer, N. J. Heinert, J. K. Doyle, and F. Yuan, "Impervious surface mapping and change monitoring using Landsat remote sensing," in ASPRS annual conference proceedings, 2004, vol. 10: American Society for Photogrammetry and Remote Sensing Bethesda, MD, USA.

[19] S. Mahyoub, H. Rhinane, M. Mansour, and A. Al Sabri, "The Use of Deep Learning in Remote Sensing for Mapping Impervious Surface: a Review Paper," in International conference of Moroccan Geomatics (Morgeo), 2020.

[20] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 152, pp. 166-177, 2019/06/01/ 2019.

[21] L. Luo, P. Li, and X. Yan, "Deep Learning-Based Building Extraction from Remote Sensing Images: A Comprehensive Review," vol. 14, no. 23, p. 7982, 2021.

[22] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," International Journal of Remote Sensing, vol. 24, no. 3, pp. 583-594, 2003/01/01 2003.

[23] H. J. P. E. Xu and R. Sensing, "Analysis of impervious surface and its impact on urban heat environment using the normalized difference impervious surface index (NDISI)," vol. 76, no. 5, pp. 557-565, 2010.

[24] C. Liu, Z. Shao, M. Chen, and H. Luo, "MNDISI: a multi-source composition index for impervious surface area estimation at the individual city scale," Remote Sensing Letters, vol. 4, no. 8, pp. 803-812, 2013/08/01 2013.

[25] Y. Tian, H. Chen, Q. Song, and K. Zheng, "A Novel Index for Impervious Surface Area Mapping: Development and Validation," vol. 10, no. 10, p. 1521, 2018.

[26] C. Deng and C. Wu, "BCI: A biophysical composition index for remote sensing of urban environments," Remote Sensing of Environment, vol. 127, pp. 247-259, 2012/12/01/ 2012.

[27] Y. Zhang, H. Zhang, and H. Lin, "Improving the impervious surface estimation with combined use of optical and SAR remote sensing images," Remote Sensing of Environment, vol. 141, pp. 155-167, 2014/02/05/ 2014.

[28] L. Shi et al., "Impervious Surface Change Mapping with an Uncertainty-Based Spatial-Temporal Consistency Model: A Case Study in Wuhan City Using Landsat Time-Series Datasets from 1987 to 2016," vol. 9, no. 11, p. 1148, 2017.

[29] J. Wang, Z. Wu, C. Wu, Z. Cao, W. Fan, and P. Tarolli, "Improving impervious surface estimation: an integrated method of classification and regression trees (CART) and linear spectral mixture analysis (LSMA) based on error analysis," GIScience & Remote Sensing, vol. 55, no. 4, pp. 583-603, 2018/07/04 2018.

[30] S. Mahyoub, H. Rhinane, A. Fadil, M. Mansour, M. Saleh, and F. Al-Nahmi, Using of Open Access remote sensing Data in Google earth engine platform for mapping built-up area in Marrakech City, Morocco. 2020, pp. 1-5.

[31] Y. J. F. Bengio and t. i. M. Learning, "Learning deep architectures for AI," vol. 2, no. 1, pp. 1-127, 2009.

[32] L. Zhang, L. Zhang, and B. Du, "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art," IEEE Geoscience and Remote Sensing Magazine, vol. 4, no. 2, pp. 22-40, 2016.

[33] X. X. Zhu et al., "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources," IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 4, pp. 8-36, 2017.

[34] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. J. R. s. o. E. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," vol. 202, pp. 18-27, 2017.

[35] P. Teluguntla et al., "A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 144, pp. 325-340, 2018/10/01/ 2018.

[36] F. Huang, Y. Yu, T. J. J. o. V. C. Feng, and I. Representation, "Automatic extraction of impervious surfaces from high resolution remote sensing images based on deep learning," vol. 58, pp. 453-461, 2019.

[37] M. A. Wulder et al., "The global Landsat archive: Status, consolidation, and direction," Remote Sensing of Environment, vol. 185, pp. 271-283, 2016/11/01/ 2016.

[38] M. A. Wulder, J. G. Masek, W. B. Cohen, T. R. Loveland, and C. E. Woodcock, "Opening the archive: How free data has enabled the science and monitoring promise of Landsat," Remote Sensing of Environment, vol. 122, pp. 2-10, 2012/07/01/ 2012.

[39] C. Deng and C. J. R. S. o. E. Wu, "A spatially adaptive spectral mixture analysis for mapping subpixel urban impervious surface distribution," vol. 133, pp. 62-70, 2013.

[40] W.-z. Yue and C.-f. J. J. O. R. S.-B.-. Wu, "Urban impervious surface distribution estimation by spectral mixture analysis," vol. 11, no. 6, p. 914, 2007.

[41] X. Zhang and S. Du, "A Linear Dirichlet Mixture Model for decomposing scenes: Application to analyzing urban functional zonings," Remote Sensing of Environment, vol. 169, pp. 37-49, 2015/11/01/ 2015.

[42] H. Ko, S. B. Hofer, B. Pichler, K. A. Buchanan, P. J. Sjöström, and T. D. J. N. Mrsic-Flogel, "Functional specificity of local synaptic connections in neocortical networks," vol. 473, no. 7345, pp. 87-91, 2011.

[43] M. Zhou, R. Zhang, W. Xie, W. Qian, and A. Zhou, "Security and Privacy in Cloud Computing: A Survey," in 2010 Sixth International Conference on Semantics, Knowledge and Grids, 2010, pp. 105-112.

[44] M. Al-Ruithe, E. Benkhelifa, K. J. P. Hameed, and U. Computing, "A systematic literature review of data governance and cloud data governance," vol. 23, no. 5, pp. 839-859, 2019.

[45] M. Belgiu, L. J. I. j. o. p. Drăguţ, and r. sensing, "Random forest in remote sensing: A review of applications and future directions," vol. 114, pp. 24-31, 2016.

[46] S. Piramanayagam, W. Schwartzkopf, F. Koehler, and E. Saber, "Classification of remote sensed images using random forests and deep learning framework," in Image and signal processing for remote sensing XXII, 2016, vol. 10004, pp. 205-212: SPIE.

E-AHP: An Enhanced Analytical Hierarchy Process Algorithm for Priotrizing Large Software Requirements Numbers

Nahla Mohamed[1], Sherif Mazen[2], Waleed Helmy[3]
Department of Information Systems, Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt

***Abstract*—One of the main activities of software requirements analysis is requirements prioritization. The wrong requirements prioritization is risky as it leads to many software failures. The current requirements prioritization techniques can't deal with large requirement numbers efficiently, which is considered one of their main issues. Many researchers have agreed that the analytical hierarchy process (AHP) is one of the best prioritization techniques as it produces highly accurate results. AHP has two main problems: scalability and inconsistency. These problems have motivated us to propose an improved version of AHP for software requirements prioritization, namely Enhanced AHP (E-AHP). A performance evaluation has been done for the conventional AHP, E-AHP, and one of the recent algorithms that also try to solve the AHP scalability problems, namely removing eigenvalues and introducing the dynamic consistency checking algorithm into AHP (ReDCCahp) algorithms The evaluation shows which algorithm takes the least time, uses the least memory, produces the most consistent and accurate results, and has the highest scalability. The three algorithms have been evaluated by running their codes using different numbers of requirements ranging from 10 to 500. The results show that E-AHP is more scalable, takes the least time, uses the least memory, and produces the most consistent and accurate results compared to the other two algorithms. That becomes remarkable when the number of requirements increases. Therefore, E-AHP is suitable to be applied in large software projects, as it can deal efficiently with the large software requirements numbers.**

Keywords*—*Requirements engineering; analytical hierarchy process; software engineering; requirements prioritization techniques

I. Introduction

Requirements engineering is a critical part of software engineering [1]. It is the process of gathering the requirements and understanding them deeply to ensure that they are correct, complete, and consistent [2]. If the requirements engineering process has not taken enough time, it will affect the overall project [3], [4]. The requirements engineering process consists of five activities: requirements elicitation, requirements analysis, requirements specification, requirements validation, and requirements management. The prioritization of requirements is one of the critical activities in the requirements analysis process [5], [6]. When the requirements number increases, analysts must organize them to implement the most important ones in the early stages to avoid the high cost of system transformation and rework and achieve user satisfaction according to a pre-specified budget, time, and resources [7].

There are three requirements prioritization technique types [8], [9]: nominal scale techniques, ordinal scale techniques and ratio scale techniques. In the Nominal scale prioritization techniques [7], [10], the users assign each requirement to a priority group, and all requirements in the same group have the same priority [8]. One of the well-known techniques is the Numerical Assignment technique, which categorizes the requirements by distributing them into groups [11]; each group has a number that describes its rank or order among all groups. And the number of groups equals the scale range [9], [12]. Top Ten Requirements is another well-known nominal scale technique. It has only one group that contains the most ten critical requirements [9]. Another technique is MoSCow, which distributes the requirements into four main groups [8]: Must-Have, Should-Have, Could-Have, and Will-Not-Have [5], [12]. These techniques are simple, easy, and fast [10]. But their results are not accurate in most cases as they don't give specific priority value to each requirement [13] and cannot deal efficiently with large requirements numbers [11].

The Ordinal scale techniques produce an ordered requirements list [10], [12], and each requirement has a specific priority [8]. They are more accurate than nominal scale techniques [9]. One of the well-known ordinal scale techniques is the Priority group [11]. It is like the Nominal scale techniques but has only three groups: High, Medium, and Low. The users prioritize and classify the requirements within the same group into another sub-group; users repeat that looping until each group has only one. Bubble sort is another well-known ordinal scale technique [11]. In this technique, the user should list the requirements and then compare every adjacent two [9]. If the second one has less priority than the first, the user swaps the order of these two requirements. The user should repeat this process for each element in the requirements list until it becomes sorted in ascending order.

Binary Search Tree (BST) is another well-known ordinal scale technique. It depends on node structure. In BST, each node represents a requirement [14] .The root node is in the first level. The last level is the ordered requirements list. BST works as follows: the user first selects one requirement to represent the top node (the root node) [8]. After that, the user iterates on the requirements list; if the requirement in the root node is more important than the requirement in the current node, the user should search in the left sub-tree to place it. Otherwise, the user searches in the right sub-tree. The user repeats this process until putting each requirement in the right place in the tree based on its priority. The ordinal scale techniques have medium scalability, consume more time, and are less easy to use than nominal scale techniques [9].

The Ratio scale techniques are similar to the Ordinal scale prioritization techniques [6], [10]. In addition, they show relative importance among all the requirements, which means they give the requirements priority values [8]. In these techniques, the users know to what extent each requirement is more important than the others [9]. Cumulative Voting (CV) is a well-known ratio scale technique that depends on the users' voting; each user has 100 points [9] and distributes the points among the requirements based on their priority [15]. Hierarchical Cumulative Voting (HCV) is a new modification of the CV technique [2]. The main difference is that HCV also prioritizes the detailed requirements (prioritizes requirements and their sub-requirements hierarchically).

Analytical Hierarchical Process (AHP) is multi-criterion decision-making and mathematical method used in many fields, including requirements prioritization [6], [9]. It selects the best decision based on pairwise comparisons among all decisions concerning many criteria [8], [13] (note that AHP will be explained in detail in section three; because the proposed algorithm is based mainly on it). It is good to use one of these techniques when the project is critical, and it is necessary to know the exact difference of importance among all the requirements [11].

Most of the prioritization techniques can’t deal with large requirements numbers and produce accurate results at the same time [2], [5], [9], [11], [13], [14], [15]. Researchers [2], [7], [9], [11], [12], [14], [15] agreed that AHP is the most accurate prioritization technique, as it is a mathematical-based method and produces highly accurate results. But they found that AHP is suitable to be used only if the requirements number is small; otherwise, it is not good as it is not scalable and sometimes suffers from inconsistency problems. Scalability means the ability of a technique to deal with a large number of requirements efficiently, and inconsistency means it sometimes produces elements that are semantically conflicting and not compatible with each other. The limitations of AHP can be summarized as follows:

- Sometimes, the results of AHP may be inconsistent because of the high human involvement [13], especially with large requirements numbers [5].
- AHP is not fast; it takes much time to work [8], [13].
- It is not easy for users to use as it needs an excellent mathematical base. It also needs time to understand how it works [13], [4].
- AHP performs n*(n-1)/2 comparisons [4], [9]where n is the number of requirements ; that means when the requirements numbers increases, the pairwise comparisons number will increase exponentially [11], which indicates it does not work well with a large number of requirements and is not scalable [3], [13].

The previous AHP limitations have motivated us to search for new ways to enhance it to deal efficiently with large requirements numbers.

The main contributions of this paper can be summarized as follows:

- Proposing a new algorithm that tries to solve the scalability problem that faces the conventional AHP and minimizes inconsistent and inaccurate results.
- An experiment that compares the proposed algorithm against the AHP and one of the best recent algorithms introduced to solve the scalability problem of AHP, namely removing eigenvalues and introducing the dynamic consistency checking algorithm into AHP (ReDCCahp), is conducted. The experiment aims to test the three algorithms' scalability, complexity, results' accuracy, and consistency.

The rest of the paper is structured as follows: Section II presents the related works on the recent techniques introduced to solve the scalability and inconsistency problems of the conventional AHP and other prioritization techniques. Section III is the research background; it explains the conventional AHP (as the proposed algorithm is a modification of AHP). Section IV presents the proposed requirements prioritization algorithm. Section V presents an experiment that compares the proposed algorithm against the AHP and ReDCCahp algorithms. Section VI presents the experimental results and discussion. Section VII is the conclusion of the paper. Section VIII is the limitations and future works.

II. Related Work

Many researchers introduced several approaches and techniques to prioritize a large number of requirements efficiently. This section will briefly explain most of them. Market-Driven Requirement Prioritization Model (MDRM) [16] is an AHP modification model introduced to deal with large requirements numbers by reducing the number of pairwise comparisons. The main idea of MDRM is to divide all the requirements into bins and prioritize all these bins by AHP. The main limitations of this technique are that it can’t consider the dependencies and conflicts among the requirements [3] and cannot deal with large requirements number efficiently [13].

NAcAHP is another technique introduced to prioritize a large number of requirements [17]; it combines the AHP technique with the Numerical Assignment technique to reduce the time that results from the pairwise comparisons [12], [11]. There are three main priority groups: Optional, Standard, and Critical. AHP works only on ones in the Critical group. One of the main limitations of this technique is that it works well only if at least 80 % of all requirements are critical (because users can't know their priorities until completing the prioritization process) [18], [19]. Other limitations [2], [3] are that it does not do consistency checking for the results, and it has not been evaluated on large data sets [17].

Fuzzy AHP [20] is another approach introduced to solve the scalability issue that faces AHP [11]. The main idea of Fuzzy AHP is to use fuzzy scales [21], and the pairwise comparison matrix consists of fuzzy triangle numbers. It provides flexibility and efficiency to get benefits from the decision-maker's preferences. This approach also addresses the uncertainty in human judgment that AHP cannot address [3], [18]. Fuzzy AHP has many limitations. One of them is that it is not reliable [8], couldn't solve the scalability problem as it can't deal with large requirements numbers and takes much time to work [13]. Another limitation [22], [23] is that it doesn't consider requirements dependencies. And also, fuzzy systems are highly dependent on human expertise, have no systematic problem-solving approach, and need a lot of validation and testing. Researchers [22] proposed a goal-based requirements prioritization technique. It depends on giving weights to the requirements based on the different project's goals [3]. But this technique is not scalable [13], suffers from the data vagueness and uncertainty problems as it heavily relies on user involvement, and does not consider the dependency relationships among the requirements [22].

The Interactive Genetic Algorithm-based (IGA) technique [24] was introduced to solve the scalability problem by combining pairwise comparisons IGA [18]. This technique uses the IGA to reduce the pairwise comparisons number [3], [14], it’s working on extracting from the user the relevant knowledge, and each user provides his preference values. IGA algorithms don’t require much information about the problem. But they have many limitations [2], [5]. Un-Scalability is a major one, as the search space increases exponentially when the number of problem elements increases [25]. Another is that the experts choose the best solution only after comparing it to the others, has no stopping criteria, and designing the objective function and getting the correct operators and representation needs effort [9].

Researchers [9] introduced an expert system, namely the Priority Handler (PHandler), to solve the scalability issue. It combines three approaches, Value-based Intelligent Requirement Prioritization (VIRP), the Back Propagation Neural Network (BPNN), and AHP. PHandler predicts the values of the requirement priorities by applying the BPNN, and then AHP. It can deal with large requirements numbers [13]. The main challenge of this system is choosing professional business analysts because a strong analyst's knowledge is necessary to estimate accurate values of requirements classification factors. One of the main limitations [13] of this system is that it neglects the dependency relationship among requirements. And the expert systems do not explain the logic behind taking a decision, cannot easily automate complex processes, and have no common sense when making a decision.

Fuzzy AHP ANN [21] is an artificial intelligence decision support system proposed to deal with large requirements numbers [18]. It integrates the Artificial intelligence Neural Network (ANN) with AHP to select the best alternatives. It determines the priority weights for the requirements using a program, namely PECAR. After that, a supervised ANN is trained (by applying a feed-forward back-propagation algorithm) using results from the PECAR program. And the decision-makers can apply different scenarios using the PECAR program by entering several input parameters into it and then observing the difference among the results. The main challenge of this system is that it needs high experts’ involvement in the prioritization process. One of the main limitations of this system [13] is that it does not produce consistent results. Another limitation is that large neural networks consume a high processing time, need a lot of data to work, cannot specify a single solution for the problem, and not scalable [2].

Researchers [26] introduced a graph-based approach to prioritize a large number of requirements. It represents the requirements as a directed graph; each node represents a requirement and can be a pre-request for or dependent on other nodes. The dependency relationships among nodes are represented as directed arrows. After that, all spanning trees are generated from the graph. In the end, requirements priorities values will be calculated based on the number of requirements dependent on them (the dependent requirements will have lower priority than the pre-requisite requirements). The main limitation of this approach [27] is the large memory consumption. It is also hard to be implemented by users; its representation is not structured, and has no specific spanning concept [26]. Researchers [4] introduced an iteration model for implementing large numbers of requirements. The main idea is to implement the requirements in phases and not all at one time. It uses the graph-based approach. One of the main limitations of this model is it does not implement all the requirements as it implements the critical ones only [4]. Another limitation is that system architecture issues will appear as all the requirements have not been gathered together, it needs more resources, and it doesn't consider the dependency among the requirements.

Researchers [19] introduced another technique based on AHP, namely, ReDCCahp. The main idea of ReDCCahp is to put every pair of adjunct requirements from the requirements list in one group and make the pairwise comparison among these groups to reduce the number of pairwise comparisons and matrix size. It is fast, simple, and does not need a strong background in math, data science, or data structure; to understand it. So it is easy to use and understand by users and more effective than AHP when dealing with a medium number of requirements. But this technique is not highly accurate as it randomly groups the adjacent requirements in the list, which means it does not have a specific requirements grouping method, which is considered its main limitation [19]. It also can deal with only small and medium requirements numbers [13] and doesn’t consider the requirements dependencies. Because ReDCCahp is one of the easiest and best new requirements prioritization techniques, the proposed algorithm tries to solve its limitation besides AHP limitations by introducing an efficient method for requirements grouping. A comparison has been made among the proposed algorithm, the conventional AHP, and the ReDCCahp algorithms.

III. Background: The Conventional AHP

This section explains the conventional AHP; because the proposed algorithm is based on it. AHP is a multi-criterion decision-making method developed by Saaty (1980) [28], [29] to solve social science domain problems. It had been used in many other fields, one of which is requirements engineering. AHP is a mathematical technique based mainly on pairwise comparisons; it selects the best decision based on comparisons among all decisions concerning many criteria. It is one of the efficient and best techniques for dealing with complex decisions. AHP consists of two phases: 1. Construct the reciprocal matrix and get the priority vector (PV). 2. Checking the consistency of the results.

In the first phase, an n × n reciprocal matrix is constructed (where n is the number of requirements) from the pairwise comparisons among the requirements by letting the user choose a value from a specified scale in AHP (each value in the scale refers to a specific importance degree) between each pair of requirements as follows: the user will put 1 in the cell(i, j) and cell(j, i) in the matrix when the requirements i and j have the same importance, and if i has more priority than j, the user should put the value he chooses from the scale in cell (i, j), and put the reciprocal of this value in cell (j, i) and vice versa. Table I shows the pairwise comparison scale in AHP.

TABLE I. PAIRWISE COMPARISON SCALE IN AHP

Intensity of importance	**Description**	**Reciprocal value**
1	Equal importance	1
2	Equal to moderate difference in importance	1/2
3	Moderate difference in importance	1/3
4	Moderate to strong difference in importance	1/4
5	Strong difference in importance	1/5
6	Strong to very strong difference in importance	1/6
7	Very strong difference in importance	1/7
8	Very strong to extremely difference in importance	1/8
9	Extreme difference in importance	1/9

Then, each element in the matrix should be divided by the sum of its columns to get a normalized matrix. After that, the user should sum elements of each row in the matrix to get the eigenvector. The last step is normalizing the eigenvector by dividing each cell by the requirements number. The normalized eigenvector is the PV, which describes the relative weights among the requirements, and the summation of all values in PV should equal one. After finishing these steps, user goes to the second phase, the Consistency checking. It means that if there are three requirements: R1, R2, and R3. R1 is more important than R2, and R2 is more important than R3, then R1 should be more important than R3. That check is called the Transitivity check. So the user should ensure that all the elements in the matrix are transitive (the matrix is a correct reciprocal matrix). The percentage of inconsistent results is high as the matrix produced by AHP is made by humans.

Professor Saaty defined a measure for consistency checking called Consistency Index (CI), which is calculated using the formula (λmax-n) / (n-1), where λmax means the maximum eigenvalue of the matrix [30], and n is the number of pairwise comparisons. To ensure that the matrix is consistent, CI should equal zero (λmax should equal n), which means there is no deviation or difference between the excepted reciprocal matrix and the resulting one. But in real-life ideal cases rarely happen, so; how much inconsistency is acceptable? Professor Saaty put a specified percentage; if the error didn't exceed it, then the matrix is consistent (there is a minimum acceptable ratio for the inconsistency).

Saaty defined this ratio as Consistency Ratio (CR), which is a guide to checking whether the matrix is consistent or not. If CR is more than 10%, then the matrix is inconsistent, and users should repeat the process from the beginning, but if the CR value is equal to or less than 10 %, then the matrix is consistent. CR value is the value of CI divided by Random Index (RI), where RI is the average CI value of several comparison matrices sizes.

IV. E-AHP: THE PROPOSED ALGORITHM

Ma [31] has found that the user's effort in the prioritization process should be reduced to solve the AHP's scalability problem. And that can happen by reducing the time pairwise comparisons take. This section introduces an AHP-based algorithm namely, Enhanced Analytical Hierarchal Process (E-AHP). Which increases the scalability of AHP by reducing the time AHP takes to construct a reciprocal matrix; its main idea is to group similar requirements using a specific method. It also decreases the inconsistent results by giving scores to the requirements groups. E-AHP consists of five main steps that will be explained in the following subsections.

A. Gathering the Requirements

First, the analysts should gather all the functional and non-functional requirements, ensure they are consistent, discover any dependencies among them [2] and make sure they are clear and specific.

B. Scoring and Sorting the Requirements

In this step, each user assigns a score to each requirement, and the score scale will equal the total requirements number. More than one requirement can have the same score if they have the same priority to the user. After finishing the scoring process, the algorithm sorts the requirements in descending order by their scores. For example, if there are three requirements: R1, R2, and R3, in this case, the score scale will be from 1 to 3. If the user assigns scores 1 to R1, 3 to R2, and 2 to R3, then the algorithm will sort them in descending order, and the sorted requirements list will be 1. R2, 2. R3, 3. R1.

To sort the requirements' scores list, a hybrid algorithm of insertion sort algorithm [19] and merge sort algorithm [32] is applied. The main idea of the hybrid algorithm is to divide the list of requirements scores into chunks. The chunk is an ascending or descending sorted sub-list that has the following patterns: $a_i > a_{i+1} > ... > a_n$ or $a_i < a_{i+1} < ... < a_n$ where a is the score of the requirement in position i, and n is requirements number. For example, if there is a list {1, 5, 6, 4, 3, 2} then the first chuck will be {1, 5, 6} (ascending order), and the second chunk will be {4, 3, 2} (descending order). After that, the algorithm reverses the first chunks to be sorted in descending order. A minimum size for each chunk is defined. For example, if the list is {2, 3, 1, 4, 5, 6} and the minimum pre-specified chunk size is 3, then the first chunk should be {2, 3, 1} not {2, 3}, although the element 1 breaks the chunk's pattern. After that, the algorithm does an insertion sort in descending order to this chunk to be {3, 2, and 1}. That algorithm fastens the sorting process (the complexity of the best case is O (n)) as it benefits from the already existing sorted sub-lists. In the end, the merge sort is applied to all these sub-lists to make a final sorted list.

C. Grouping the Similar Requirements

After sorting the requirements list, the algorithm will group all requirements that have the same score, or the difference among their scores is <= MaxDRS. Where the MaxDRS variable is the Maximum Difference of Requirements Scores in the same group; and it is pre-specified by the user. The user also specifies the value of MaxNR, which is the Maximum Number of Requirements in one group. For example, if there are five requirements: R1, R2, R3, R4, and R5. R1 and R2 have a score of 1, R3 has a score of 2, R4 has a score of 5, R5 has a score of 6, and MaxDRS = 1. Then the algorithm will put R1, R2, and R3 in one group and R4; and R5 in another group. And if MaxNR = 5 and seven of them have a score <= the specified MaxDRS, in this case, only five of them will be in the same group, and the rest two will be in another new group. To assign a score for one group, the algorithm calculates the average score of all its requirements. For example, if a group has three requirements: R1, R2, and R3, and the score of R1, R2 is 1, and the score of R3 is 2, then the score of their group will be (1+1+2)/3 =1.3.

The different values of MaxNR and MaxDRS influence the results, because there are negative relationships between the values of MaxNR, MaxDRS, and time, and between them and accuracy. If users aim to decrease the time taken, the MaxNR and MaxDRS values should be increased, which reduces the accuracy and vice versa. The best choice is to choose the values of MaxNR and MaxDRS based on the total number of requirements in the project. If the requirements number is large, it's better to choose large values for them and vice versa.

D. Constructing the Reciprocal Matrix

In these steps, E-AHP constructs the reciprocal matrix like AHP, but in E-AHP, the rows and columns of the matrix are the requirements groups, and the matrix's elements are the difference between the scores of the requirements groups. For example, if the score of G1 is 4; and the score of G2 is 2, then the element in the intersection of row G1, column G2 will be 2. And will be -2 between row G2 and column G1. After that, E-AHP normalizes the elements in the matrix and does the same mathematical calculations in AHP to get the PV. The flow chart in Fig. 1 and Algorithm I explain the steps to get the PV in E-AHP.

Algorithm I. Construct PC matrix and get the PV

```
Input: N - number of requirements
Input: RN - the requirements names list
Input: RS - the requirements scores list
Input: MaxDRS - maximum difference of requirements
scores in one group
Input: MaxNR - maximum numbers of requirements in one
group
Output: List of prioritized requirements
Initialize: R      // list of requirements objects
Initialize: G      // One group of requirements
Initialize: GL     // One group length
Initialize: AG     // List of all groups
Initialize: NG     // Number of all groups
Initialize: Sum    // Counter
Initialize: M      // The Reciprocal matrix
Initialize: CS     // List of column sum
Begin
for i in N do
    requirement ← new Requirement()
    requirement.name ←RN[i]
    requirement.score ← RS[i]
    requirement.ingroup ← false
    R.append(requirement)
end for
R ← sort(R by names, reverse=true)
for i in R.length do
    if R[i].ingroup == true then
        skip to next iteration
    G.reqList.append(R[i])
    R[i].ingroup ← true // R[i] is assigned to a group
    for j in R.length do
        if R[j].ingroup == true
        or (R[j].score - R[i].score) > MaxDRS
        or G.length > MaxNR then
            skip to next iteration
        else
            G.reqList.append(R[i])
            R[i].ingroup ← true
        end if
        G.reqList ← []
        G.avScore ← 0.0
    end for
    AG.append(G)
end for
NG ← AG.length
for i in NG do
    if AG[i].length == 1 then
        AG[i].avScore←AG[i][0].score
    else
        GL ← AG[i].length
        for j in GL do
            AG[i].avScore ←AG[i].avScore +
            AG[i].reqList[j].score
        end for
        AG[i].avScore←AG[i].avScore / GL
    end if
end for
for i in NG do
    for j in NG do
        if AG[i].avScore == AG[j].avScore then
            M[i][j] ← 1
        else if AG[i].avScore > AG[j].avScore then
            M[i][j]←AG[i].avScore - AG[j].avScore
        else
            M[i][j] ←1 / (AG[i].avScore -
            AG[j].avScore)
        end if
    end for
end for
Sum ←0
for j in NG do
    for i in NG do
        Sum ← Sum + M[i][j]
    end for
```

```
        CS[j] ←Sum
        Sum ← 0
    end for
    for j in NG do
        for i in NG do
            M[i][j] ← M[i][j] / CS[j]
        end for
        Sum ←0
    end for
    for i in NG do
        PV[i] ←PV[i] / NG
    end for
    for i in NG do
        for j in AG[i]
            print(AG[i].reqList[j].name "has relative weight
            : " + PV[i])
        end for
    end for
    End
```

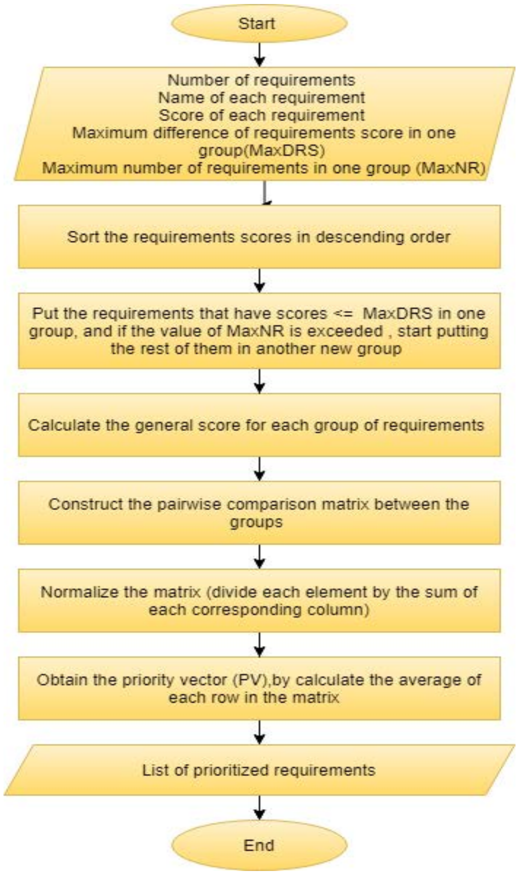

Start
Number of requirements
Name of each requirement
Score of each requirement
Maximum difference of requirements score in one group(MaxDRS)
Maximum number of requirements in one group (MaxNR)
Sort the requirements scores in descending order
Put the requirements that have scores <= MaxDRS in one group, and if the value of MaxNR is exceeded , start putting the rest of them in another new group
Calculate the general score for each group of requirements
Construct the pairwise comparison matrix between the groups
Normalize the matrix (divide each element by the sum of each corresponding column)
Obtain the priority vector (PV),by calculate the average of each row in the matrix
List of prioritized requirements
End

Fig. 1. A Flowchart shows Steps to Get the PV in the E-AHP Algorithm.

E. Consistency Check of the Results

In this step, a new consistency check algorithm is applied to ensure that the results are consistent. This step has three inputs: X, Y, and Z. X is a list of the most critical requirements to the user, and Y is the number of first groups from the resulting PV that will be checked. And Z is the minimum acceptable number of requirements in the list X that must be in the Y groups. The user fills the list (X), and to prove consistency, the algorithm checks that at least (Z) of them are in the (Y) groups from the resulting PV; otherwise, the result is inconsistent. And users should repeat the prioritization process from the beginning and re-evaluate their preference judgments. For example, if the size of the list X =5, Y=6 and Z=3, which means the user will choose the most critical five requirements for him, and then to prove the consistency of the results, the algorithm must find at least three of these requirements in the first sex groups from the resulting PV. These variables influence the results as the more the values of X, Z and the less the value of Y, the more accurate the results will be.

The flow chart in Fig. 2 and Algorithm II explain the consistency check step in E-AHP.

Algorithm II. Consistency checking of the results

```
Input: X - list of most important requirements
Input: Y - number of requirements for check
Input: Z – minimum accepted number of found requirements
Input: PV - priority vector result from phase 1
Output: print whether the results are consistent or not
Initialize: XL // length of most important requirements list
Initialize: nXF // number of important requirements found
Begin
nXF ← 0
for i in XL do
    for j in Y do
        if X[i] == PV[j] then
            nXF ← nXF + 1
        end if
    end for
end for
if nXF >= Z
    print ( "Results are consistent")
else print ("Results are not consistent")
end if
End
```

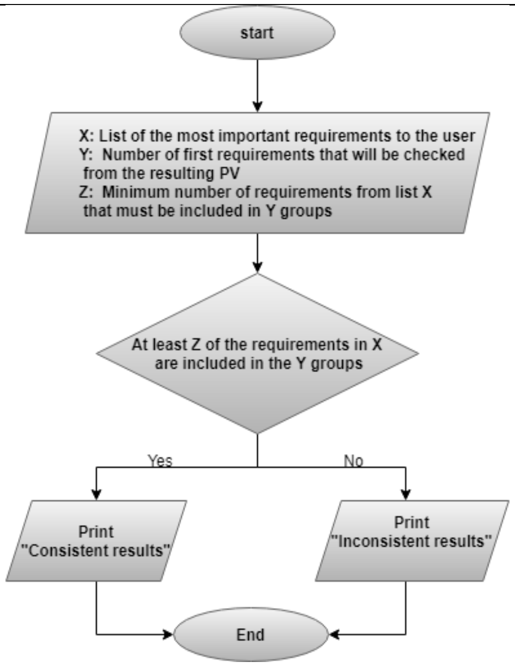

start
X: List of the most important requirements to the user
Y: Number of first requirements that will be checked from the resulting PV
Z: Minimum number of requirements from list X that must be included in Y groups
At least Z of the requirements in X are included in the Y groups
Yes
No
Print "Consistent results"
Print "Inconsistent results"
End

Fig. 2. A Flowchart shows the Consistency Check Step in the E-AHP Algorithm.

V. An Experiment Comparing the Proposed Algorithm (E-AHP) Against AHP and ReDCCahp Algorithms

This section will explain the experiment that compares the E-AHP against the AHP and ReDCCahp. It will mention the experiment objectives, variables, and setup.

A. *Experimental Objective*

This experiment aims to validate that E-AHP (the proposed algorithm) is better than AHP and ReDCCahp; by proving that it solves the scalability, inconsistency, and accuracy problems (as AHP suffers from scalability and inconsistency problems, and ReDCCahp suffers from accuracy problems). The experiment has compared the three algorithms (AHP, ReDCCahp, and E-AHP) against each other to test which one can deal with large numbers of requirements efficiently.

The research aims to answer the following questions:

- Which algorithm among the three algorithms takes the least time?
- Which algorithm among the three algorithms uses the least memory?
- Which algorithm among the three algorithms produces the most consistent results (produces results with minimum CR value)?
- Which algorithm among the three algorithms is the most scalable?
- Which algorithm among the three algorithms produces the most accurate results?

B. *Experimental Variables*

The experiment has three independent variables: AHP, ReDCCahp, and E-AHP, and three dependent variables: time, memory, and CR value. A brief definition for these dependent variables is given below.

- Time: representing the time each algorithm takes to prioritize the requirements (completion time of each algorithm), and it is measured in minutes.
- Memory: representing the memory needed for each algorithm to prioritize the requirements, and it is measured in megabytes (MB).
- CR: is the main criterion to check whether the results are consistent. It is calculated in the experiment by the same equation used in AHP (CI / RI) [28], [29].

C. *Experimental Setup*

The time consumption, memory usage, and CR values are evaluated by implementing and running the algorithms' codes and comparing their results. They have been written in the JAVA programming language and run on a machine with a Processor: Intel(R) Core(TM) i7-6500U CPU @ 2.50 GHz 2.60 GHz, and Installed Memory (RAM): 8.00 GB and System type: 64-bit operating system, x64-based processor.

Researchers [9], [31] considered the requirements numbers small when they are less than 20; medium when they are between 20 and 50; Otherwise, large. Several data set sizes (small, medium, and large) that range from 10 to 500 requirements are used in the experiment to prove the efficiency of E-AHP over AHP and ReDCCahp when dealing with various requirements numbers. The input will be a list of requirements objects (R1, R2, and R3…Rn), where n is the requirements number. Each requirement object has the name and score of the requirement. The requirements names are according to their order in the list. To consider the different scores for the requirements, the codes of the three algorithms have run ten times, each with different scores values, and then the average results are taken. Only the requirements numbers and their scores are important in the experiment, and it doesn't matter about their meaning.

VI. Discussion and Results

A. *Discussion*

This section presents and discusses the experiment results; it compares the performance of AHP, ReDCCahp, and E-AHP. It will show the time taken, memory used, and CR values of the algorithms after running their Java code with different numbers of requirements ranging from 10 to 500. Tables II and III present the average time taken and memory used by AHP, ReDDCahp, and E-AHP, respectively. Table IV presents the CR values of the AHP, ReDCCahp, and E-AHP results.

TABLE II. THE AVERAGE TIME TAKEN (IN MINUTES) BY THE AHP, REDDCAHP AND E-AHP ALGORITHMS

Number of requirements	**AHP**	**ReDDCahp**	**E-AHP**
10	2.501	1.391	0.297
25	9.647	3.675	0.972
50	17.533	6.988	2.326
100	27.091	13.962	4.955
150	49.081	22.491	6.883
200	74.718	28.541	13.481
300	109.981	46.846	25.236
400	136.345	61.932	36.766
500	161.709	95.961	45.152

TABLE III. THE AVERAGE MEMORY USED (IN MB) BY THE AHP, REDDCAHP AND E-AHP ALGORITHMS

Number of requirements	**AHP**	**ReDDCahp**	**E-AHP**
10	0.0681	0.0344	0.0293
25	0.2133	0.0939	0.0845
50	0.3312	0.1997	0.1497
100	0.5508	0.3138	0.2349
150	0.7019	0.4609	0.2911
200	0.9402	0.5011	0.3278
300	1.3082	0.7443	0.4026
400	1.8708	0.9952	0.5133
500	2.8937	1.1960	0.6479

TABLE IV. THE AVERAGE CR VALUE OF THE AHP, REDCCAHP AND E-AHP ALGORITHMS' RESULTS

Number of requirements	AHP	ReDDCahp	E-AHP
10	0.0791	0.0504	0.0193
25	0.1944	0.0905	0.0337
50	0.2908	0.0981	0.0617
100	0.3873	0.1649	0.0729
150	0.4521	0.2591	0.0801
200	0.5091	0.3287	0.0896
300	0.6099	0.4926	0.11091
400	0.7011	0.6608	0.1482
500	0.8019	0.7492	0.1615

Charts in Fig. 3 and Fig. 4 visualize results in Table II and Table III, respectively. Chart in Fig. 5 visualizes results in Table IV. Tables II, III, Fig.3, and Fig. 4 show that ReDCCahp and E-AHP take less time and memory than AHP. When the requirements number is from 10 to 50, the difference between the time taken and memory used by the algorithms is small, larger when the number of requirements becomes > 50, and significant when > 150. That happens because they apply the grouping method, which decreases their matrix size (when the matrix size decreases, the memory and time needed to complete the operations to get the final PV decreases).

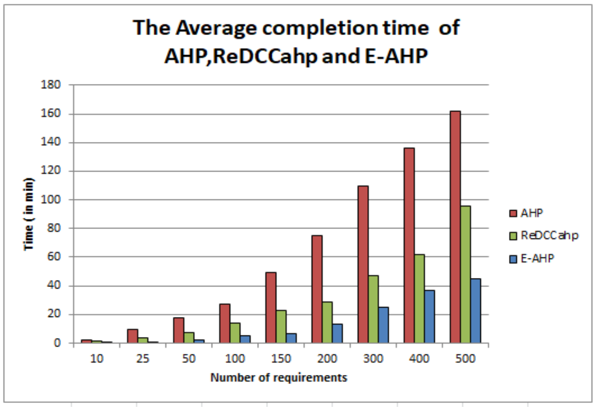


Fig. 3. The Average Time taken by the AHP, ReDCCahp and E-AHP Algorithms.

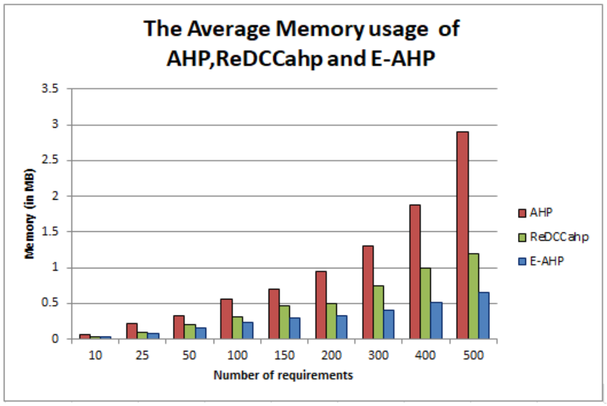


Fig. 4. The Average Memory used by the AHP, ReDCCahp and E-AHP Algorithms.

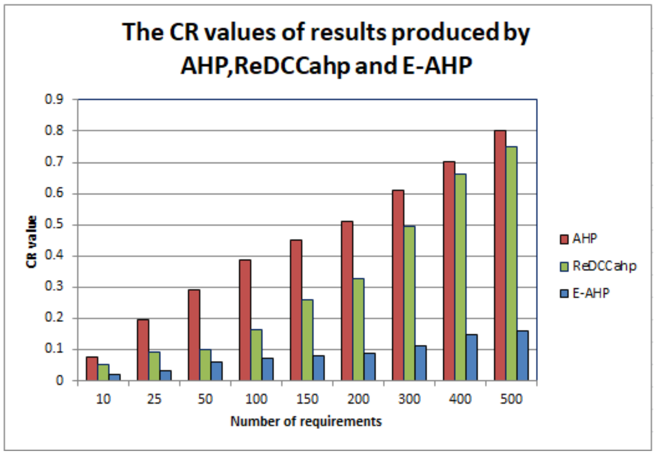


Fig. 5. The Average CR Value of the AHP, ReDCCahp and E-AHP Algorithms' Result.

It also can be noticed that E-AHP takes less time and memory than ReDCCahp, especially when the requirements number is > 100 (large). Because, in ReDCCahp, the group size is fixed (each group has only two requirements), but in E-AHP, the group size is variable (one group can have any requirements number). So in most cases, E-AHP produces less number of groups than ReDCCahp, which makes its pairwise comparison matrix size smaller than ReDCCahp.

Table IV and Fig. 5 show that the results of AHP becomes inconsistent (CR > .1) when the number of requirements is > = 25. That happens because it requires the users to make pairwise comparisons among all requirements; not among requirements groups, which increases the human error proportion; and hence decreases the results consistency. ReDCCahp produces consistent results when the number of requirements is < =50, and E-AHP produces consistent results when the requirement number is < 300. Although E-AHP and ReDCCahp both group the requirements, E-AHP produces more consistent results than ReDCCahp. That happens because E-AHP uses the scoring method instead of the pairwise comparisons, which decreases the human error proportion and increases results consistency. Moreover, the scoring method takes less time than the pairwise comparisons method, as the pairwise comparisons increase the time exponentially when the requirements number increases.

B. Results

The results that can be concluded from the experiment are as follows:

- AHP consumes more time than ReDCCahp and E-AHP, and E-AHP uses the least time among the three algorithms.
- AHP uses more memory than ReDCCahp and E-AHP, and E-AHP uses the least among the three algorithms.
- AHP produces less consistent results than ReDCCahp and E-AHP, and E-AHP produces the most consistent results among the three algorithms.
- AHP has less scalability than ReDCCahp and E-AHP, and E-AHP is the most scalable algorithm among the three algorithms.

- AHP produces less accurate results than ReDCCahp and E-AHP, and E-AHP produces the most accurate results among the three algorithms.

So among the three algorithms, E-AHP is the best one as it takes the least time, uses the least memory, has the highest scalability, and produces the most consistent and accurate results. All of that becomes remarkable when the number of requirements increases.

VII. Conclusion

This research proposes a new software requirements prioritization algorithm to solve the scalability and inconsistency problems faced by the AHP, namely E-AHP. A performance evaluation of the E-AHP algorithm against the AHP and ReDCCahp algorithms (ReDCCahp is one of the best recent algorithms that try to solve the AHP problems) was done to prove the effectiveness and efficiency of E-AHP. The java codes of the three algorithms have been implemented and run on the same machine with various requirements numbers ranging from 10 to 500 (small, medium, and large). The time taken, the memory used, and CR values of the results are measured. Results show that E-AHP takes much less time, uses less memory, and produces more consistent and accurate results than AHP and ReDCCahp, especially with large requirements numbers, which means that E-AHP has high scalability. So E-AHP is better than AHP and ReDCCahp as it can deal efficiently with large numbers of requirements.

VIII. Limitations and Future Work

Some cases will reduce the speed and accuracy of the E-AHP algorithm: first, if the difference among most of the requirements scores is more than the MaxDRS value, in this case, most of the groups will have one requirement, and the number of groups will increase (will almost equal to the requirements number), which will cause a scalability problem. Second, if the difference among most requirements scores is the same, then each group will have many requirements, which will decrease the accuracy of the results. The negative effect is reduced in these cases by choosing small values for MaxDRS and MaxNR. So in the future, we plan to enhance the E-AHP algorithm to deal with the previous cases efficiently. We also plan to conduct an experiment using a large number of software analysts as participants to validate the applicability and usability of the proposed algorithm on large real-life software projects.

References

[1] Rashidah Kasauli, Eric Knauss, Jennifer Horkoff, Grischa Liebel, Francisco Gomes de Oliveira Neto,"Requirements engineering challenges and practices in large-scale agile system development", Journal of Systems and Software, 2021.

[2] Naila Jan, Irum Inayat, and Muhammad Abbas, "An Empirical Evaluation of Requirements Prioritization Techniques." , Marketing and Branding Research,2020.

[3] Faiza Allah Bukhs , Zaharah Allah Bukhsh , and Maya Daneva, "A systematic literature review on requirement prioritization techniques and their empirical evaluation", Computer Standards,2020.

[4] Muhammad Yaseen, Noraini Ibrahim, and Aida Mustapha, "Requirements prioritization and using iteration model for successful implementation of requirements", International Journal of Advanced Computer Science and Applications, 2019.

[5] Khaled AbdElazim, Ramadan Moawad, and Essam Elfakharany, "A framework for requirements prioritization process in agile software development", Journal of Physics: Conference Series, 2020.

[6] Emmanuel OC Mkpojiogu, and Nor Laily Hashim, "Quality based prioritization: An approach for prioritizing software requirements", Journal of Telecommunication, Electronic and Computer Engineering, 2018.

[7] Hanny Tufail, Iqra Qasim, Muhammad Faisal Masood, Sara Tanvir, Wasi Haider Butt, "Towards the selection of Optimum Requirements Prioritization Technique: A Comparative Analysis.", 2019 5th International Conference on Information Management (ICIM). IEEE, 2019.

[8] Iroju Olaronke, Rhoda Ikono, Ishaya Gambo, "An Appraisal of Software Requirement Prioritization Techniques", Asian Journal of Research in Computer Science, 2018.

[9] Muhammad Imran Babar, Masitah Ghazali, Dayang N.A. Jawawi, Siti Maryam Shamsuddin, and Noraini Ibrahim, "PHandler: an expert system for a scalable software requirements prioritization process", Knowledge-Based Systems, 2015.

[10] Jamilah Din, Muhammed Basheer Jasser, "Software Requirements Prioritization Tool using a Hybrid Technique ", International Journal of Engineering and Advanced Technology (IJEAT), 2019.

[11] Philip Achimugu, Ali Selamat, Roliana Ibrahim, and Mohd Nazri Mahrin, "A systematic literature review of software requirements prioritization research", Information and software technology, 2014.

[12] Nayak, Soumen, Chiranjeev Kumar, and Sachin Tripathi. "Analytic hierarchy process-based regression test case prioritization technique enhancing the fault detection rate.", Soft Computing 26.15, 2022.

[13] Fadhl Hujainah, Rohani Binti Abu Bakar, Mansoor Abdullateef Abdulgabber,and Kamal Z. Zamli, "Software requirements prioritisation: a systematic literature review on significance, stakeholders, techniques and challenges.", IEEE Access ,2018.

[14] Noor Hazlini Borhan, Hazura Zulzalil, Sa'adah Hassan, Norhayati Mohd Ali,"Requirements Prioritization Techniques Focusing on Agile Software Development:A Systematic Literature review ", International Journal of Scientific and Technology Research, 2019.

[15] Amjad Hudaib, Raja M.T Masadeh, Mais Qasem, and Abdullah Issa Alzaqebah, "Requirements prioritization techniques comparison." Modern Applied Science, 2018.

[16] Muhammad Atif Iqbal, Athar Mohsin Zaidi, and Saeed Murtaza, "A new requirement prioritization model for market driven products using analytical hierarchical process", 2010 International Conference on Data Storage and Data Engineering .IEEE, 2010.

[17] Srinivas Nidhra, Likith Poovanna, Kelapanda Satish, and Vinay Sudha Ethiraj, "Analytical Hierarchy Process issues and mitigation strategy for large number of requirements", 2012 CSI Sixth International Conference on Software Engineering (CONSEG). IEEE, 2012.

[18] Naila Jan, Irum Inayat,and Muhammad Abbas, ", An Empirical Evaluation of Requirements Prioritization Techniques ,Marketing and Branding Research", 2020.

[19] Iyas Ibriwesh, Sin-Ban Ho, and Ian Chai, "Overcoming scalability issues in analytic hierarchy process with ReDCCahp: An empirical investigation", Arabian Journal for Science and Engineering, 2018.

[20] Xiaojun Wang, Hing Kai Chan, Rachel W.Y. Yee , and Ivan Diaz-Rainey, "A two-stage fuzzy-AHP model for risk assessment of implementing green initiatives in the fashion supply chain", International Journal of Production Economics, 2012.

[21] Yash Veer Singh, Bijendra Kumar, Satish Chand, and Jitendra Kumar ,"A comparative analysis and proposing 'ANN fuzzy AHP model'for requirements prioritization.", I.J. Information Technology and Computer Science, 2018.

[22] Mukhtar Elsood, A. Abo, Hesham A. Hefny, and Eman S. Nasr, "A goal-based technique for requirements prioritization", 2014 9th International Conference on Informatics and Systems. IEEE, 2014.

[23] YanLiu, Claudia M.Eckert,and ChristopherEarl. "A review of fuzzy AHP methods for decision-making with subjective judgements", Expert Systems with Applications, 2020.

[24] Paolo Tonella, Angelo Susi, and Francis Palma, "Using interactive GA for requirements prioritization", 2nd International Symposium on Search Based Software Engineering. IEEE, 2013.

[25] Sourabh Katoch, Sumit Singh Chauhan, and Vijay Kumar, "A review on genetic algorithm: past, present, and future", Multimedia Tools and Applications, 2021.

[26] Muhammad Yaseen , Noraini Ibrahim , Aida Mustapha, "Prioritization of Software Functional Requirements: Spanning Tree based Approach ",International Journal of Advanced Computer Science and Applications,2019.

[27] Lafore, Robert. Data structures and algorithms in Java. Sams publishing, 2017.

[28] Thomas L. Saaty, "What is the analytic hierarchy process? ", Mathematical models for decision support.Springer, Berlin, Heidelberg, 1988.

[29] Bruce L. Golden, Edward A. Wasil, and Patrick T. Harker, "The analytic hierarchy process", Applications and Studies, Berlin, Heidelberg, 1989.

[30] Chatelin, Françoise, ed. Eigenvalues of matrices: revised edition. Society for Industrial and Applied Mathematics, 2012.

[31] Qiao. Ma, The effectiveness of requirements prioritization techniques for a medium to large number of requirements: a systematic literature review", Diss. Auckland University of Technology, 2009.

[32] Kurt. Mehlhorn, "Data structures and algorithms 1: Sorting and searching", Springer Science & Business Media, 2013.

Effectiveness of Human-in-the-Loop Sentiment Polarization with Few Corrected Labels

Ruhaila Maskat[1]*, Nurzety Aqtar Ahmad Azuan[2], Siti Auni Amaram[3], Nur Hayatin[4]
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia[1, 3]
Faculty of Computing, College of Computing & Applied Sciences, Universiti Malaysia Pahang[2]
Informatics Department, University of Muhammadiyah Malang, Jawa Timur, Indonesia[4]

***Abstract*—In this work, we investigated the effectiveness of adopting Human-in-the-Loop (HITL) aimed to correct automatically generated labels from existing scoring models, e.g. SentiWordNet and Vader to enhance prediction accuracy. Recently, many proposals showed a trend in utilizing these models to label data by assuming that the labels produced are near to ground truth. However, none investigated the correctness of this notion. Therefore, this paper fills this gap. Bad labels result in bad predictions, hence hypothetically, by positioning a human in the computing loop to correct inaccurate labels accuracy performance can be improved. As it is infeasible to expect a human to correct a multitude of labels, we set out to answer the questions of "What is the smallest percentage of corrected labels needed to improve prediction quality against a baseline?" and "Would randomly selecting automatic labels for correction produce better prediction than specifically choosing labels with distinct data points?". Naïve Bayes (NB) and Decision Tree (DT) were employed on AirBnB and Vaccines public datasets. We could conclude from our results that not all ML algorithms are suited to be used in a HITL environment. NB fared better than DT at producing improved accuracy with small percentages of corrected labels, as low as 1%, exceeding the baseline. When selected for human correction, labels with distinct data points assisted in enhancing the accuracy better than random selection for NB across both datasets, yet partially for DT.**

Keywords*—Human-in-the-loop; few labels; sentiment polarization

I. Introduction

Sentiment is used in an array of applications, from sentiment analysis to customer intelligence. Two primary sentiment polarities are "positive" and "negative". In the absence of any sentiment, a "neutral" polarity would be given. To date, additional polarities have been introduced to include the element of intensity. They are "strongly positive" and "strongly negative". Typically, the approach to determining polarity is via a hand-crafted lexicon of sentiments. Words in the lexicon were earlier identified to express positive or negative opinions about a particular subject of interest. This approach is limited to the list of collected words. Continuous enrichment must occur to sustain an ever-expanding vocabulary, especially on social media platforms with the existence of new words e.g., "Google" and "tweet". On its own, this approach may not be perpetually effective when more data are added.

Evolving from this approach is the use of machine learning (ML) algorithms trained on a dataset labelled with polarity. Through this approach, the algorithms learn from the labels and predict the polarity of newly unseen data points. This frees arduous efforts in the upkeeping of a lexicon. However, the performance of this approach relies considerably on a large number of good-quality labels which usually are produced with the assistance of human annotators. Several downsides of this technique are human annotators can be scarce in some domains, the quality of the annotators can differ depending on their level of knowledge and experience, not all datasets can be annotated, engaging human annotators are costly and they are incapable of annotating a tremendously large number of data points.

Recent works [1]–[5] show the implementation of a hybrid solution where a lexicon is used to automatically label a dataset which is then used as a training set by a ML algorithm. The assumption made is the produced labels are close to the ground truth, thus reliable. Our work investigates this assumption based on two lexicon-based scoring models of different types – a valence-based lexicon, SentiWordNet, and a rule-based lexicon, Vader. Human annotators were employed to examine the generated sentiment polarity of two public datasets (AirBnB and Vaccines) when the resulting labels conflict with one another. We found that Vader identifies polarity more like a human annotator than SentiWordNet. We extended this finding to further explore the effect Human-in-the-Loop (HITL) has on accuracy. By having a human expert in the loop, incorrect labels that compromised quality can be emended. Consequently, the ML algorithm learns the correction and updates its knowledge space, resulting in enhanced accuracy. As human annotators would not be able to correct a multitude of labels, we added another component to this experimental setting, i.e., few labels, which is counterintuitive to many ML algorithms as they often need numerous labelled data.

In this work, we determined the quality of sentiment labels from SentiWordNet and Vader based on a human's opinion. We constructed a HITL experimental framework and addressed two important questions. 1) What is the smallest percentage of corrected labels needed to improve prediction quality against a baseline? 2) Would randomly selecting automatic labels for correction produce better prediction than specifically choosing automatic labels with distinct data points?

This paper is outlined as follows. The literature review is covered in Section II. In Section III, we describe our proposed HITL methodology and formulated two research questions to be addressed via experiments. The results of the experiments are presented and discussed in Section IV. Finally, in Section V, we conclude this paper.

II. Literature Review

A. Sentiment Labelling

Traditional sentiment labelling requires human annotators to label data. Although this is known to produce gold standard datasets, unfortunately, it can be error-prone, time-consuming, labour intensive and infeasible with big data [6]. Automatic labelling comes into the picture to overcome this limitation. As of late, a scoring model such as SentiWordNet [7] or Vader [6] has been employed for this task. SentiWordNet works by utilizing a lexicon of synonymous English words clustered together a.k.a. synset. It contains 147,306 synsets labelled with the polarity of positivity, negativity and neutrality. An advantage of SentiWordNet as compared to older lexicons, e.g., Linguistic Inquiry and Word Count (LIWC), is its account of sentiment valence to assist in portraying sentiment intensity [6]. Nonetheless, SentiWordNet also shares the disadvantages of other lexicons: a shortage in coverage, where essential features tend to be missed and costly maintenance [6]. Conversely, Vader is newer and leverages rules to decide on the polarity and intensity of sentiments [6]. Vader adopted the same human-validated sentiment lexicons as LIWC, Affective Norms for English Words (ANEW) and General Inquirer (GI) yet appended more features specific to social media text, making it sensitive to their nuances. The cornerstone of Vader is the adoption of the wisdom-of-the-crowd to determine sentiment valence and the leveraging of Grounded Theory to formulate rules customized to generalize across an array of grammatical and syntactical functions for sentiment polarity decision-making.

B. Human-in-the-Loop (HITL)

HITL is not a new idea. Including feedback from a human during a computer-related process to improve effectiveness has been earlier proposed – e.g., harnessing paid feedback in crowdsourcing via Amazon Turk [8]. Another example is in the Pay-as-You-Go dataspace where users supply feedback to assist in resolving entities during data integration [9]. The HITL idea is also adopted in Few-Shot Learning [10], Active Learning [11], Transfer Learning [12] and User Guidance [13]. The underlying notion is performance can be enhanced even with a small amount of data from humans, especially for datasets that are too large to the extent of being infeasible to manually annotate [14].

C. Machine Learning Algorithms

Naïve Bayes and Decision Tree are popularly used in numerous sentiment analysis literature. To date, they include [1], [15]–[19].

1) Naïve Bayes (NB) when used for sentiment prediction is simple and intuitive yet highly accurate [20]. To predict, NB leverages the concept of conditional probability. Its advantage is it does not require a large training set [21]. However, a primary disadvantage is its tendency to theorize linguistic features to be independent under soconditionsion [20]. The argument put forth is the nature of words is co-occurring and are joined by syntactic as well as semantic dependencies, hence, NB may produce unfavourable performance.

2) Decision Tree (DT) is hierarchical in its natural form. Rules are generated for the task of predicting target terms. Each leaf node contains a word feature of the sentences in a corpus. DT was reported to have a great deal of adaptability to large datasets as compared to other ML algorithms [21]. A larger dataset triggers the formulation of additional rules, permitting the construction of higher quality trees with a larger pool of attributes available [22]. Nonetheless, this can be disadvantageous in a setting with few labels.

III. Research Methodology

A. Automatic Sentiment Production

In this step, we produced labels on each data point using automatic techniques. Two popularly used techniques are SentiWordNet and Vader scoring models. We applied both techniques to AirBnB[1] and Vaccines[2] public datasets. AirBnB contains 142,114 reviews on various premise owners in Asheville, North Carolina, United States, while the Vaccines dataset consists of 24,075 tweets related to opinions of Covid-19 vaccines worldwide. Both techniques' output is a pair of polarity confidence values for each data point; one for positive sentiment and the other for negative sentiment. The higher confidence value between the two becomes the final suggested sentiment. E.g., the positive polarity confidence for data *X* is 25% whereas the negative polarity confidence is 75%. Therefore, data point *X* will have a negative polarity.

When the set of sentiment labels is placed in tandem, three sets of results exist. The first is where both models scored positive sentiments on each data point. The second set contains only negative sentiments, and the last set has contradicting sentiments. For example, Vader scored a positive sentiment whereas SentiWordNet scored a negative sentiment on the same data point *Y*. The earlier two sets are not interesting in this study since both techniques produced the same result. We assume there is strong evidence to support the generated degree of confidence which led to the same result and thus it is more worthwhile to bring our focus to the contradicting set as this would allow us to discriminate between the two scoring models. For AirBnB, the size of the contrasting set was 3,967. For the Vaccines dataset, a total of 3,055 tweets were found.

B. Human Labelling

Next thing is to determine which contradicting labels are correct, Vader's or SentiWordNet's. Known as the ground truth, this task must be performed by the participation of a human. We employed two human annotators to evaluate each suggested sentiment in the contradicting set. The annotators are not English native speakers but have more than five years of writing and speaking English at a university level. We aim to investigate which model is less aligned with a human's

[1] http://insideairbnb.com/get-the-data.html

[2] https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets

judgement. This model is then used to discover the role of HITL with few corrected labels in improving the performance of prediction.

The result in Table I displays Vader produced 85.73% labels that match with human labels from the AirBnB dataset. From the Vaccines dataset, 22.37% of Vader's labels match with human labels. In contrast, for the AirBnB dataset, SentiWordNet obtained 2.14% matches and 4.45% for the Vaccines dataset. In summary, SentiWordNet scored the farthest to a human's perception of sentiment on both datasets. Additionally, for the conflicting set, we discovered that SentiWordNet tends to score text as negative sentiment more than positive, whereas Vader was the opposite. This could be due to Vader's capability to handle social media features as the datasets contain online reviews of that nature. Furthermore, both the datasets are general, and thus it would be interesting to use an inherently negative dataset in the future, for example from the mental health domain, to see how both models would behave. Therefore, with these two datasets, Vader produced better scores compared to SentiWordNet.

TABLE I. SCORING MODELS MATCH WITH HUMAN LABEL

AirBnB		***Vaccines***	
Vader	**SentiWordNet**	**Vader**	**SentiWordNet**
85.73%	2.14%	**22.39%**	4.45%

C. Calculate Baseline Effectiveness

Since SentiWordNet and Vader show contradictory results to human labels, we then proceed to calculate baseline effectiveness i.e. the accuracy that could be achieved. Fig. 1 shows the processes involved. Each label set from SentiWordNet and Vader was used to train two popularly used ML algorithms, Decision Tree and Naïve Bayes. Cross-validation of five folds was employed. Afterwards, predictions of each dataset were generated. To know the accuracy of these predictions, we compare them against human labels.

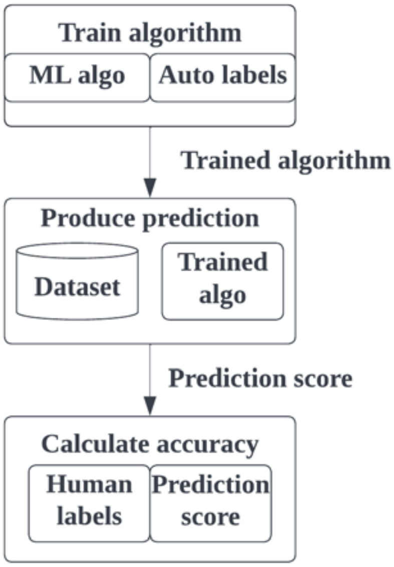


Fig. 1. Process of Calculating Baseline Effectiveness.

TABLE II. ACCURACY VALUES OF BOTH SCORING MODELS WITH SENTIWORDNET CHOSEN AS THE BASELINE

	Decision Tree (DT)		***Naïve Bayes (NB)***	
	AirBnB	**Vaccines**	**AirBnB**	**Vaccines**
SWordNet	**1.34%**	**0.00%**	**0.00%**	**0.00%**
Vader	86.54%	26.84%	87.88%	26.84%

Table II displays the baseline accuracy values. SentiWordNet labels produced substantially low accuracy across both datasets and both ML algorithms with values of 0.00% and 1.34%. Conversely, Vader labels achieved good accuracy for the AirBnB dataset of 86.54% (DT) and 87.88% (NB); the average for the Vaccines dataset with 26.84% (both DT and NB). To conclude, SentiWordNet's low-quality labels are less effective in a prediction task.

From the findings, it is interesting to learn the effects on prediction accuracy when a human expert is present to provide explicit feedback in the form of corrections to automatic labels. Hence, we performed several experiments to investigate.

D. Experimental HITL Framework Construction

To conduct this experiment, we developed a basic HITL framework and adapted it to incorporate human explicit feedback (Fig. 2). Our study did not lean towards any specific variants of HITL, alternatively, our interest is in exploring the generic idea of including humans in the labelling process to correct a small number of labels and observe the outcome. The framework consists of five layers: data, automatic label generation, algorithm training, prediction performance evaluation and label correction.

In the first layer, datasets are introduced into the framework. From these datasets, the scoring models will automatically generate labels. These scoring models are in-built with their own lexicon and thus do not require any annotated data point. Then, a human expert checks a small number of the produced labels and corrects them, if deemed necessary. By choosing to correct the labels or otherwise, the expert inadvertently injects explicit feedback into the framework. The output, a set of labels, becomes the training set for one or more machine learning algorithms. Once trained, the algorithms will produce predictions. The quality of these predictions, in the form of accuracy, is calculated to measure the effectiveness of the corrected automatic labels. A threshold of preferred accuracy is checked and if this threshold was not reached, the predicted labels are presented to the human expert for correction. Several iterations would occur until the threshold is met. Alternatively, if the human expert decided not to continue correcting anymore labels, the loop would end. This manifests the role of a human in the prediction loop; thus, the term human-in-the-loop framework.

The following questions are addressed in this work and the answers are explained in the next section.

Q1: What is the smallest percentage of corrected labels needed to improve prediction quality against a baseline?

Q2: Would randomly selecting automatic labels for correction produce better prediction than specifically choosing automatic labels with distinct data points?

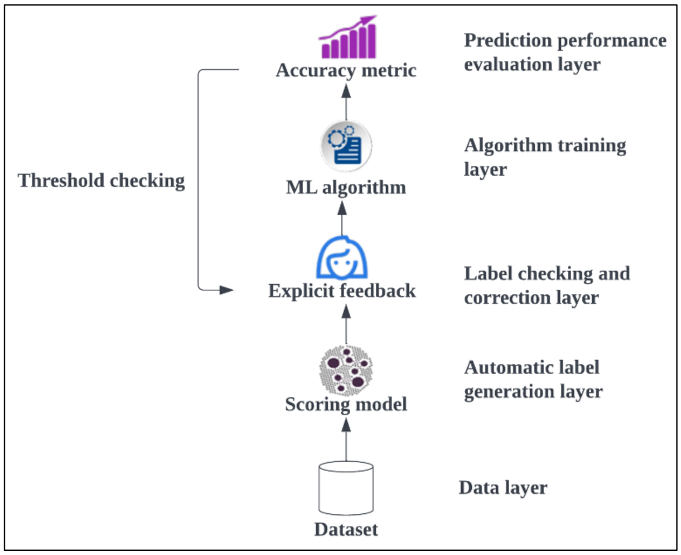


Fig. 2. Experimental HITL Framework.

IV. Result and Discussion

In this section, we describe the experimental setup and present as well as discuss the results and findings of our study. RapidMiner and Orange were employed as the simulation platform for these experiments.

A. *Experiment 1*

Aim: This experiment aims to answer Q1 where we want to determine at what percentage of corrected labels introduced in the loop would enhance accuracy. The threshold accuracy values are based on the baseline results in Table II. Since SentiWordNet yields accuracy values worse than Vader, therefore, it was used as the scoring model in this study.

Setup: We experimented with different percentages of human labels. Very small percentages of 10% and below were tested, followed by percentages of 20% and 35%. The labels were randomly selected. Two classic ML algorithms were used, Decision Tree (DT) and Naïve Bayes (NB). Cross validation of 5 folds was employed.

Result: Referring to Table III and Fig. 3, we found the following:

1) *Baseline test:* Only NB exceeded the baseline accuracy for both datasets (NB AirBnB – 99.14; NB Vaccines – 69.56). DT surpassed the baseline accuracy for the Vaccines dataset at 73.16, however, failed for AirBnB, reaching the highest accuracy of only 1.34 which is at par with the baseline.

2) *Small percentages test:* NB was able to achieve improvement in accuracy at even 1% of corrected labels for both AirBnB and Vaccines datasets. In contrast, DT did not improve for AirBnB but showed marked improvement (i.e. 73.16) only when 35% of corrected labels were supplied.

In summary, not all ML algorithms are suitable for HITL with few corrected labels. Nevertheless, with the right algorithm, a percentage of corrected labels as small as 1% can be effective in significantly enhancing accuracy. Depending on the characteristics of the dataset, further increment of accuracy beyond the initial value can occur early as witnessed from AirBnB when NB is employed. A jump in accuracy of 46.63 from 12.13 was attained at 9%. To note, these labels were chosen randomly, therefore, an interesting alternative is where a more deliberate strategy is tested. Experiment 2 explores this idea.

B. *Experiment 2*

This experiment investigates if by carefully choosing automatic labels with distinct data points would yield better accuracy than when the labels were picked randomly. The result of this experiment answers Q2. Distinct data points represent unique cases within the data space of a particular domain. This approach trains ML algorithms using a set of "small data" [23]. Small data supports the notion of quality over quantity. Here, a small number of high-quality data points that represents the majority of a population is more preferred than a large, primarily uniformed, collection of data points. Such uniformity can cause the algorithm to become blindsided and thus focuses only on a specific case, limiting its learning experience.

TABLE III. SentiWordNet Accuracy Values

	Decision Tree (DT)		***Naïve Bayes (NB)***	
	AirBnB	**Vaccines**	**AirBnB**	**Vaccines**
Baseline	1.34%	0.00%	0.00%	0.00%
1%	0.91%	0.00%	**12.12%**	**69.20%**
2%	1.01%	0.00%	12.12%	69.20%
3%	**1.34%**	0.00%	12.13%	69.20%
4%	1.08%	0.00%	12.13%	69.20%
5%	0.91%	0.00%	12.12%	69.20%
6%	1.34%	0.00%	12.13%	69.20%
7%	1.34%	0.00%	12.13%	69.20%
8%	1.34%	0.00%	12.13%	69.20%
9%	1.34%	0.00%	**46.63%**	69.20%
10%	0.66%	0.00%	81.46%	69.20%
20%	1.34%	0.00%	98.66%	69.20%
35%	1.34%	**73.16%**	**99.14%**	**69.56%**

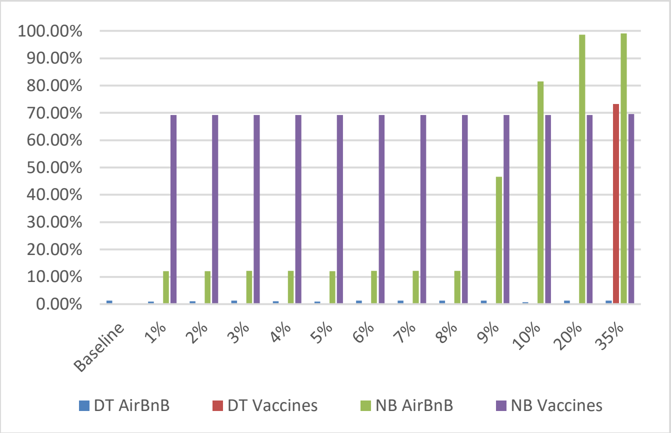


Fig. 3. SentiWordNet Accuracy Values Visualized with Bars.

To reflect the distinct nature of the data points, four techniques to calculate the distance between a pair of text were used and compared. They are Cosine, Euclidean, Jaccard and Manhattan. Afterwards, the resulting similar texts were clustered together. A dendrogram was formed and a cutting point was determined based on the production of a cluster set with a size of approximately 30 to 40 clusters. The rationale behind this condition is to produce an approximate minimum number of data points as in Experiment 1 for reasons of comparison fairness. In other words, 1% of AirBnB and Vaccines datasets, each. These data points and their labels were used to train both DT and NB. Stratification was included in this experiment to understand its possible influence on effectively producing better accuracy when there are very few human-corrected labels. Therefore, a combined total of 8 techniques were used. They are Non-Stratified Cosine (NSC), Non-Stratified Euclidean (NSE), Non-Stratified Jaccard (NSJ), Non-Stratified Manhattan (NSM), Stratified Cosine (SC), Stratified Euclidean (SE), Stratified Jaccard (SJ) and Stratified Manhattan (SM).

Setup: Alike Experiment 1, we experimented with different percentages of human labels. Very small percentages of 10% and below were tested, followed by percentages of 20% and 35%. The labels were obtained from across the derived cluster set to consist of as many unique representations as possible. Two classic ML algorithms were used: Decision Tree (DT) and Naïve Bayes (NB). Cross validation of 5 folds was employed.

Result: The following were discovered.

1) Baseline test: Generally, the results exhibit a similar pattern as found in Experiment 1, but with enhanced accuracy values in a majority of the cases. Table IV and Fig. 4 show the average accuracy of all eight techniques when no stratification was used with distinct data points. The result shows no difference in the accuracy of AirBnB when applied to DT where in both random and distinct data, DT did not surpass the baseline. In contrast, for Vaccine dataset, a higher accuracy was obtained at 78.00 while random data achieved only 73.16. With NB, it attained accuracy higher than the baseline for both datasets (NB AirBnB – 99.77; NB Vaccines – 77.79) and better than random data in Experiment 1.

When we compare to see if stratification of data can contribute to the improvement of accuracy, we found that this is true for NB but not for DT. Accuracy for AirBnB remained similarly low as the baseline i.e. 1.34 in DT. In addition, DT achieved only a slightly higher accuracy i.e. 78.07 with Vaccines dataset as compared to random data i.e. 78.00. Conversely, NB reached better accuracy in AirBnB and Vaccines datasets with stratification (NB AirBnB – 99.94; NB Vaccines – 77.85). Table V and Fig. 5 show the average accuracy for distinct data points with stratification applied.

Thus far, we can observe that in this HITL framework, taking advantage of distinct data points can produce better accuracy than just randomly selecting data for a human expert to check and correct. Additionally, applying stratification can further improve the produced accuracy. To understand the performance of each of the eight techniques, we selected the best values they produced against the baseline and the result is displayed in Table VI.

TABLE IV. AVERAGE ACCURACY FOR DISTINCT DATA POINTS WITHOUT STRATIFICATION APPLIED

	Decision Tree (DT)		*Naïve Bayes (NB)*	
	AirBnB	**Vaccines**	**AirBnB**	**Vaccines**
Baseline	1.34%	0.00%	0.00%	0.00%
1%	1.25%	0.00%	**12.12%**	**70.59%**
2%	1.26%	0.00%	12.13%	71.40%
3%	1.26%	0.00%	12.13%	71.87%
4%	**1.34%**	0.00%	12.13%	72.50%
5%	1.28%	0.00%	12.13%	72.96%
6%	1.34%	0.00%	12.13%	73.30%
7%	1.26%	0.00%	12.13%	73.37%
8%	1.34%	0.00%	**20.79%**	73.64%
9%	1.34%	0.00%	46.77%	73.76%
10%	1.34%	0.00%	94.32%	73.81%
20%	1.34%	**34.28%**	98.66%	75.19%
35%	1.34%	**78.00%**	**99.77%**	**77.79%**

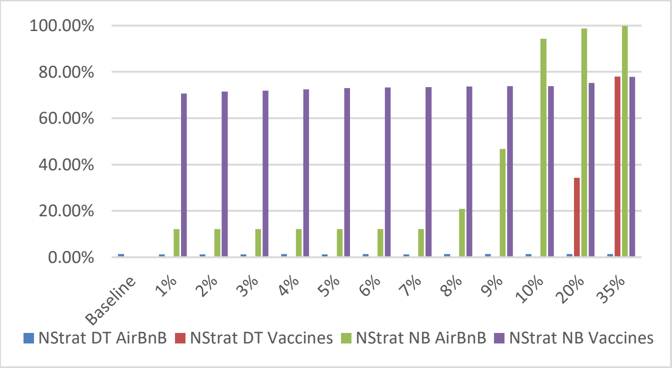


Fig. 4. Average Accuracy for Non-stratified Data Visualized with Bars.

TABLE V. AVERAGE ACCURACY FOR DISTINCT DATA POINTS WITH STRATIFICATION APPLIED

	Decision Tree (DT)		*Naïve Bayes (NB)*	
	AirBnB	**Vaccines**	**AirBnB**	**Vaccines**
Baseline	1.34	0.00	0.00	0.00
1%	**1.34**	0.00	**12.12**	**70.01**
2%	1.22	0.00	12.12	70.59
3%	1.28	0.00	12.13	70.82
4%	1.28	0.00	12.13	70.97
5%	1.34	0.00	12.13	71.76
6%	1.34	0.00	12.13	72.39
7%	1.34	0.00	12.13	72.48
8%	1.34	0.00	**22.45**	72.89
9%	1.34	0.00	74.33	73.09
10%	1.34	0.00	98.66	73.16
20%	1.34	**30.48**	98.66	75.36
35%	1.34	**78.07**	**99.94**	**77.85**

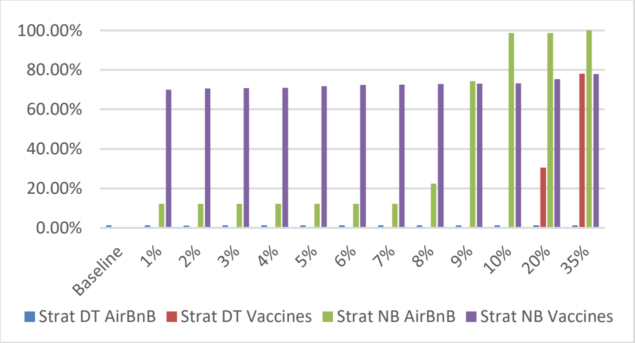


Fig. 5. Average Accuracy for Stratified Data Visualized with Bars.

TABLE VI. BASELINE TEST RESULT

	Decision Tree (DT)		***Naïve Bayes (NB)***	
	AirBnB	**Vaccines**	**AirBnB**	**Vaccines**
NSC	1.34%	78.54%	99.77%	78.33%
NSE	1.34%	77.66%	99.60%	77.44%
NSJ	1.34%	**78.88%**	99.72%	**78.68%**
NSM	1.34%	76.93%	**100.00%**	76.70%
SC	1.34%	78.01%	**100.00%**	77.80%
SE	1.34%	77.77%	**100.00%**	77.55%
SJ	1.34%	78.09%	**100.00%**	77.88%
SM	1.34%	75.90%	98.66%	75.67%

NSC-Non-Stratified Cosine; NSE-Non-Stratified Euclidean; NSJ-Non-Stratified Jaccard; NSM-Non-Stratified Manhattan; SC-Stratified Cosine; SJ-Stratified Euclidean; SJ-Stratified Jaccard; SM-Stratified Manhattan

None of the eight techniques successfully excelled over the baseline when DT was applied to the AirBnB dataset, but when NB was applied all surpassed the baseline for both datasets (Table VI). Furthermore, the top values from all the techniques were also generated when NB was used with AirBnB, even reaching 100% accuracy. For the Vaccines dataset, both DT and NB produced comparable accuracy values between 75 and 79, indicating generalization occurred well here. Jaccard, despite having stratification or otherwise, continually churned good accuracy in both AirBnB and Vaccines datasets (DT Vaccines – 78.88; NB AirBnB – 100; NB Vaccines – 78.68), but performed better with stratification.

2) Small percentages test: Similar to the result in Experiment 1, DT did not produce better accuracy for Vaccines dataset at 1% corrected labels. Nevertheless, using distinct data points resulted in the need of a smaller percentage i.e. 20% as compared to random selection i.e. 35%. Furthermore, better accuracy was obtained in this experiment i.e., 78.00 (Table IV) from 73.16 in Experiment 1 (Table III) with the same percentage of corrected label of 35%.

As with Experiment 1, NB showed better accuracy from baseline even at 1% of corrected labels for both datasets (Table IV). For AirBnB, the accuracy at 1% is identical to Experiment 1 (i.e., 12.12), yet, the improvement of accuracy occurred faster at 8% in contrast to 9% in Experiment 1. Unfortunately, this came with a cost of reduced accuracy i.e. 20.79 (Table IV). Applying stratification did not result in needing a smaller percentage different from without stratification.

In summary, supplying corrected labels with distinct data points can help in obtaining higher accuracy if coupled with a ML algorithm suitable for HITL as proven with NB on AirBnB and Vaccines datasets. Although both have different characteristics, indicating a generalized effect, more datasets will need to be tested to ascertain this. Overall, the role of stratification in affecting accuracy was positive, depending on the combination of a ML algorithm and dataset chosen. Applying an ensemble of similarity techniques can yield better result.

V. CONCLUSION

In this paper, we have investigated the effectiveness of using human-in-the-loop (HITL) to improve prediction accuracy by correcting automatically generated labels from existing scoring models such as SentiWordNet and Vader. As more recent work adopted the use of these scoring models in place of a training set, we took the initiative to understand if their inherent assumption of these labels being gold-standard-worthy is plausible. We experimented using two public datasets, AirBnB and Vaccines, in combination with two ML algorithms, Naïve Bayes and Decision Tree, where we discovered that Naïve Bayes produced better accuracy than Decision Tree at small percentages of corrected human labels. We also discovered that selecting labels with distinct data points to be corrected helps to enhance accuracy for Naïve Bayes but partially for Decision Tree.

ACKNOWLEDGMENT

The Malaysian government is funding this research through the Fundamental Research Grant Scheme (FRGS) at Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia (FRGS/1/2019/SS05/UITM/02/5).

REFERENCES

[1] M. Aufar, R. Andreswari, and D. Pramesti, "Sentiment Analysis on Youtube Social Media Using Decision Tree and Random Forest Algorithm: A Case Study," 2020 International Conference on Data Science and Its Applications, ICoDSA 2020, 2020, doi: 10.1109/ICoDSA50139.2020.9213078.

[2] E. H. Almansor, F. K. Hussain, and O. K. Hussain, "Supervised ensemble sentiment-based framework to measure chatbot quality of services," Computing, vol. 103, no. 3, pp. 491–507, 2021, doi: 10.1007/s00607-020-00863-0.

[3] J. P. Pinto and V. M. T., "Real Time Sentiment Analysis of Political Twitter Data Using Machine Learning Approach," International Research Journal of Innovations in Engineering and Technology (IRJIET), vol. 6, no. 4, pp. 4124–4129, 2019, [Online]. Available: www.irjet.net.

[4] A. Borg and M. Boldt, "Using VADER sentiment and SVM for predicting customer response sentiment," Expert Systems with Applications, vol. 162, p. 113746, 2020, doi: 10.1016/j.eswa.2020.113746.

[5] M. R. A. Rahim, Y. Mahmud, and S. Abdul-Rahman, "Customers' Opinions on Mobile Telecommunication Services in Malaysia using Sentiment Analysis," International Journal of Advanced Computer Science and Applications, vol. 12, no. 12, pp. 222–227, 2021, doi: 10.14569/IJACSA.2021.0121229.

[6] E. Hutto, C.J. and Gilbert, "VADER: A Parsimonious Rule-based Model for," Eighth International AAAI Conference on Weblogs and Social Media, p. 18, 2014, [Online]. Available: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109.

[7] S. Baccianella, A. Esuli, and F. Sebastiani, “SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining,” Proceedings of the 7th Conference on Language Resources and Evaluation LREC10, pp. 417–422, 2008, [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf.

[8] J. Wang, S. Oyama, H. Kashima, and M. Kurihara, “Learning an accurate entity resolution model from crowdsourced labels,” Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2014, pp. 0–7, 2014, doi: 10.1145/2557977.2558060.

[9] R. Maskat, N. W. Paton, and S. M. Embury, Pay-as-you-go configuration of entity resolution, vol. 10120. 2016.

[10] K. Bailey and S. Chopra, “Few-Shot Text Classification with Pre-Trained Word Embeddings and a Human in the Loop,” pp. 1–8, 2018, [Online]. Available: http://arxiv.org/abs/1804.02063.

[11] L. Toumanidis, P. Kasnesis, C. Chatzigeorgiou, M. Feidakis, and C. Patrikakis, “ActiveCrowds: A human-in-the-loop machine learning framework,” Frontiers in Artificial Intelligence and Applications, vol. 338, pp. V–VI, 2021, doi: 10.3233/FAIA210090.

[12] L. Yang, S. Hanneke, and J. Carbonell, “A Theory of Transfer Learning with Applications to Active Learning,” Machine learning, vol. 90, no. 2, pp. 161–189, 2013.

[13] T. T. Nguyen, M. Weidlich, H. Yin, B. Zheng, Q. V. Hung Nguyen, and B. Stantic, “User guidance for efficient fact checking,” Proceedings of the VLDB Endowment, vol. 12, no. 8, pp. 850–863, 2018, doi: 10.14778/3324301.3324303.

[14] X. Wu, L. Xiao, Y. Sun, J. Zhang, T. Ma, and L. He, “A Survey of Human-in-the-loop for Machine Learning,” 2021, [Online]. Available: http://arxiv.org/abs/2108.00941.

[15] B. Lakshmi DeviV., V. Bai, and K. Somula Ramasubbareddy Govinda, “Sentiment Analysis on Movie Reviews,” Emerging Research in Data Engineering Systems and Computer Communications, pp. 321–328, 2020.

[16] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, “The impact of features extraction on the sentiment analysis,” Procedia Computer Science, vol. 152, pp. 341–348, 2019, doi: 10.1016/j.procs.2019.05.008.

[17] M. Guia, R. R. Silva, and J. Bernardino, “Comparison of Naive Bayes, support vector machine, decision trees and random forest on sentiment analysis,” IC3K 2019 - Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, vol. 1, no. Ic3k, pp. 525–531, 2019, doi: 10.5220/0008364105250531.

[18] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, “Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes,” 2018 International Conference on Orange Technologies, ICOT 2018, no. October, 2018, doi: 10.1109/ICOT.2018.8705796.

[19] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm,” Procedia Computer Science, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.

[20] P. Gamallo and M. Garcia, “Citius: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets,” 8th International Workshop on Semantic Evaluation, SemEval 2014 - co-located with the 25th International Conference on Computational Linguistics, COLING 2014, Proceedings, no. SemEval, pp. 171–175, 2014, doi: 10.3115/v1/s14-2026.

[21] J. Singh, G. Singh, and R. Singh, “Optimization of sentiment analysis using machine learning classifiers,” Human-centric Computing and Information Sciences, vol. 7, no. 1, 2017, doi: 10.1186/s13673-017-0116-3.

[22] M. Bilal, H. Israr, M. Shahid, and A. Khan, “Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques,” Journal of King Saud University - Computer and Information Sciences, vol. 28, no. 3, pp. 330–344, 2016, doi: 10.1016/j.jksuci.2015.11.003.

[23] J. J. Faraway and N. H. Augustin, “When small data beats big data,” Statistics and Probability Letters, vol. 136, pp. 142–145, 2018, doi: 10.1016/j.spl.2018.02.031.

Fuzzy Clustering Analysis of Power Incomplete Data based on Improved IVAEGAN Model

Yutian Hong*, Jun Lin
Guangdong Electric Power Information Technology Co., Ltd.
Guangzhou, Guangdong 520000 China

***Abstract*—The scale of data generated by the complex and huge power system during operation is also very large. With the data acquisition of various information systems, it is easy to form the situation of incomplete power data information, which cannot guarantee the efficiency and quality of work, and reduce the security and reliability of the entire power grid. When incomplete data and incomplete data sets are caused by data storage failure or data acquisition errors, fuzzy clustering of data will face great difficulties. The fuzzy clustering of incomplete data of the power equipment is divided into the processing of incomplete data and the clustering analysis of "recovered" complete data. This paper proposes an IVAEGAN-IFCM interval fuzzy clustering algorithm, which uses interval data sets to fill in the incomplete data, and then completes the clustering of interval data. At the same time, the whole numerical data set is transformed into a complete interval data set. The final clustering result is obtained by interval fuzzy mean clustering analysis of the whole interval data set. Finally, the algorithm proposed in this paper and other machine learning training data sets is made for experimental analysis. The experimental results show that the algorithm proposed in this paper can complete incomplete data sets with high precision clustering. Compared with other contrast methods, it shows higher clustering accuracy. Compared with the numerical clustering algorithm, the clustering accuracy is improved by more than 4.3%, and it has better robustness. It also shows better generalization on the artificial data sets and other complex data sets. It is helpful to improve the technical level of the existing power grid and has important theoretical research value and engineering practice significance.**

Keywords—Power system; power equipment; incomplete data; fuzzy clustering; mining algorithm

I. Introduction

With the development of the economy, along with the construction of the smart grid and smart distribution network informatization, the level of automation and interactive are raised. The IoT degree of mutual penetration and integration of power distribution networks as a direct link between the user and the power grid is also improved. The safe and stable operation of the distribution network is directly related to the electricity quality and reliability of power use. The reliability of the distribution network system is particularly important [1]. Due to the massive increase of all kinds of power data and the increase of complexity, the data processing speed and processing capacity are greatly improved with the help of computer clustering analysis, and it is widely used in many key parts of power [2].

Distribution network equipment will produce a large amount of data in operation, because in the whole distribution system, there are many equipment from different manufacturers. Their models are mostly different because these distribution equipment and background monitoring system will use their own transmission protocol communication, resulting in the existence of many communication methods. Due to the conversion of various interfaces and transmission protocols, it is inevitable to cause problems such as slow data transmission rate and loss of some data information. If these missing data are not effectively analyzed, the maintenance efficiency will be low, and the equipment operation status cannot be timely, accurately and finely evaluated. Finally, it will affect the operation quality and efficiency of the large power grid.

With the in-depth study of incomplete data attribute imputation algorithm, it is further found that interval data samples can better express the ambiguity of incomplete data and improve the accuracy and robustness of clustering [3]. In the fields of pattern recognition, image processing, fuzzy control, and clustering analysis, there were some relevant researches [4]. Scholars in China and abroad have proposed many improved algorithms and applications for interval fuzzy sets. Zhang and other scholars proposed an improved BP neural network interval filling incomplete data clustering algorithm, which uses a neural network to fill the interval range of missing attributes [5]. Li et al. proposed an interval kernel function fuzzy clustering algorithm, which uses the nearest neighbor rule to determine the missing attribute interval, and uses the kernel function fuzzy C-means to map and cluster the samples in high dimensions [6]. The above research content has obvious advantages for dynamic data and big data, but it is prone to complex and difficult computing problems in model construction. Trang and other scholars proposed an interval fuzzy co-clustering algorithm, which applies interval data to co-clustering to make the clustering results more accurate [7]. Reference [8] uses mutually exclusive maintenance, simultaneous maintenance, power grid security, resources and other constraints to constrain interrelated equipment, which comprehensively considers the status of power grid equipment, and power grid operation, and uses a tabu search algorithm to optimize the starting period of power transmission and transformation equipment maintenance in the whole network. The calculation methods in references [7] and [8] are relatively simple, but the accuracy of estimation needs to be further improved. The above research content does not fully tap the potential value of incomplete data, and the effective information is not fully utilized. The data processing of distribution network equipment operation big data will mainly

*Corresponding Author.

study the incomplete data objects in the distribution network equipment operation big data. The vector data repair technology based on genetic nearest neighbor clustering will study the data repair technology-oriented to the distribution network operation big data to facilitate the filling and repair of incomplete distribution network operation data objects [9].

To solve the problem of data loss in large-scale data, a deep learning generation model based on the conditional generation antagonistic network is designed according to the data characteristics. An optimization function is proposed according to the data characteristics. The replacement of training data is optimized, and the optimization constraints are proposed to make the interpolation model more effective. Based on the IVAEGAN-FCM algorithm, this paper proposes an IVAEGAN-IFCM interval fuzzy clustering algorithm, which uses interval data sets to cluster incomplete data. This study is closely related to the operation status of power equipment, which is conducive to improving the safe operation efficiency of power grid enterprises.

The main innovations of this paper are as follows:

1) A certain estimation error exists in the estimation of the missing attribute through the IVAEGAN model. The average value of all errors is taken as the interval size range of the interval type data to construct an interval type data set.

2) Calculate the extreme value of the attribute of the nearest neighbor sample of the incomplete data to restrict the interval size of each data.

3) Calculate the local density of the sample in the adjacent area of each sample, and further constrain the size of the interval through the local density of the sample.

4) Convert the whole numerical data set into a complete interval data set. To improve the accuracy of incomplete data clustering, the interval fuzzy C-means (IFCM) clustering analysis is carried out on the whole interval data set.

The main contents of this paper are as follows:

1) This paper introduces the importance of equipment data integrity in the power system operation.

2) In the related work, the power system equipment data and the IVAEGAN model are introduced.

3) The construction of the IVAEGAN interval model and the calculation of data integrity by the fuzzy clustering algorithm are done.

4) The IFCM algorithm is improved by using the IVAEGAN model.

5) The accuracy and effectiveness of the proposed algorithm are verified by experimental simulation analysis.

6) Finally, the research contents and results of this paper are summarized.

II. Related work

A. Analysis of Data Sources for Typical Power Equipment

Power grid automation helps power grid dispatchers to grasp the operating conditions of the power grid under their jurisdiction in real-time so that dispatchers can make correct dispatching decisions. At the same time, it can also provide data support such as load forecast for short-term and medium-term power grid production and development plans [10]. The centralized control mode of the substation is established to realize more and more automation systems with different functions. To improve the reliability, efficiency, and power supply quality of distribution network operation, the distribution automation system has been developed. At present, the dispatching automation system, substation integrated automation system, and distribution automation system have become the main components of the power grid automation system [11]. The power data studied in this paper takes the distribution equipment data as an example, and carries out data resources according to the equipment material information, as shown in Table I.

B. IVAEGAN Value Estimation Principle

1) An IVAEGAN model performs estimated value filling on missing attributes x_{ik} in an incomplete data set. Since the IVAEGAN model also performs estimated value calculation on data with complete attributes in a training process, an absolute error exists between an expected estimated value and an actual value, and the absolute value of an error average value is used as the interval size of the estimated value of the missing attributes. The left and right endpoints of the valuation interval are: x_{ik}^{-} , x_{ik}^{+} respectively. Then the incomplete data can be expressed in the form of interval, namely $\left[x_{ik}^{-}, x_{ik}^{+}\right]$;

For the complete data x in the incomplete data set, its complete attribute value x_{ij} also needs to be converted into the form of interval data, that is $\left[x_{ik}^{-}, x_{ik}^{+}\right]$, where $x_{ij}^{-} = x_{ij} = x_{ij}^{+}$. Therefore, the complete attribute of the complete data is also expressed in the interval type, and the left and right interval endpoints of the interval attribute are equal to the attribute value [12].

TABLE I. Power Cable Index System Table

part	First class indicator	Secondary indicator (state)
power cable	Cable body	Line load, Insulation resistance, Exterior, Fire fire flame retardant, Depth, filthy
	Cable terminal	
	Cable middle head	
	Cable channel	
	Auxiliary facilities	

Through the training process of the IVAEGAN and the evaluation process of incomplete data, the average evaluation error of the training process is determined as the range of interval size. The evaluation of incomplete data attributes is determined as the median of the interval to determine the evaluation interval of the incomplete data [13]. And then the complete numerical data set after the valuation "recovery" is transformed into a complete interval data set. Finally, the complete interval data set is subjected to fuzzy clustering analysis through the IFCM clustering algorithm, and the clustering result is obtained [14].

III. IVAEGAN Fuzzy Clustering Algorithm for Interval Estimation

A. IVAEGAN Model Construction

VAE (Variational autoencoder) is fused as the generator of GAN (Generative Adversarial Network), and the fused model is improved to obtain the IVAEGAN model. The IVAEGAN network is proposed to better estimate, predict and fill the missing data attribute values contained in the incomplete data set [15]. The core idea of the IVAEGAN network proposed in this paper is to feedback the difference between the predicted value of the incomplete data through the network generator and the discriminator to the input layer so that the network can obtain more information, and thus can better realize the estimation filling of missing attributes in incomplete data. This can make the filling value of missing attributes more reasonable, thus improving the effectiveness of clustering analysis. The topology of the improved model is shown in Fig. 1:

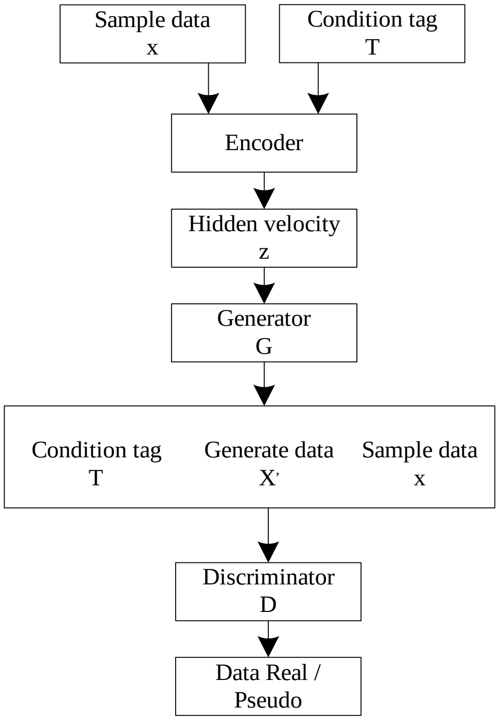


Fig. 1. IVAEGAN Model Structure.

The integration of variational learning and adversarial learning provides strong support for learning and reasoning in generative models, and these methods are used to establish new hybrid reasoning methods.

B. Interval reconstruction of numerical incomplete data set

Interval data $x=[x^-,x^+]$ and interval data $y=[y^-,y^+]$: the addition operation of interval type is expressed as: $x+y=\left[x^-+y^-,x^++y^+\right]$. Similarly, the definition of subtraction operation of interval data is expressed as: $x-y=\left[x^--y^-,x^+-y^+\right]$ [16]. Euclidean distance is commonly used in calculating relative distance, and the Euclidean distance formula for interval data is:

$$D(x,y)=\sqrt{(x^--y^-)^2+(x^+-y^+)^2} \tag{1}$$

The density of samples in a local area reflects the degree of clustering of samples, and also reflects the similarity between samples. The greater the density of samples in a local area, the closer the attribute values are. In this paper, the regional density of sample points is used to calculate sample density a, and the obtained attribute estimation interval $[\min,\max]$ is constrained to obtain a new interval: $\left[\min+\frac{a(\max-\min)}{2},\max-\frac{a(\max-\min)}{2}\right]$.

The calculation formula for the distance between samples x_p and other samples is shown in (2):

$$d_{pq}=\left\|x_p-x_q\right\|_2^2 \tag{2}$$

Let the attribute dimension be the s-interval dataset $\overline{X}=\{\overline{x}_1,\overline{x}_2,\cdots,\overline{x}_n\}$. The data $\overline{x}_i=[\overline{x}_{1i},\overline{x}_{2i},\cdots,\overline{x}_{si}]^T$, for any $\overline{x}_{ji}=\left[x_{ji}^-,x_{ji}^+\right]$, the objective function formula of the interval fuzzy C-means algorithm is shown in (3):

$$J_m(U,\overline{V})=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m\left\|\overline{x}_k-\overline{v}_i\right\|_2^2 \tag{5}$$

$$\sum_{i=1}^{c}u_{ik}=1,k=1,2,\cdots,n \tag{4}$$

$$\left\|x_k-v_i\right\|_2^2=\sqrt{(x_k{}^--v_i{}^-)^T(x_k{}^--v_i{}^-)+(x_k{}^+-v_i{}^+)^T(x_k{}^+-v_i{}^+)} \tag{5}$$

$\overline{v}_i$ represents the ith cluster center. $\overline{V}$ is the cluster center matrix, and is expressed as: $\overline{V}=\left[\overline{v}_{ji}\right]=[\overline{v}_1,\overline{v}_2,\cdots,\overline{v}_c]$, $\overline{v}_{ji}=\left[v_{ji}^-,v_{ji}^+\right]$, $i=1,2,\cdots,c$, $j=1,2,\cdots,s$ 。

The minimum condition of Equation (5) is:

$$v_i{}^-=\frac{\sum_{k=1}^{n}u_{ik}^m x_k^-}{\sum_{k=1}^{n}u_{ik}^m},i=1,2,\cdots,c \tag{6}$$

$$v_i^+ = \frac{\sum_{k=1}^{n} u_{ik}^m x_k^+}{\sum_{j=1}^{n} u_{ik}^m}, i = 1, 2, \cdots, c \tag{7}$$

If there is an interval type data sample $\overline{x}_k$ within the interval value of a cluster center, its membership degree is set to 1; otherwise, its membership degree is 0 and it does not belong to this category. The formula is as follows:

$$u_{ij} = \begin{cases} 0, i \neq h \\ 1, i = h \end{cases} \tag{8}$$

Otherwise：

$$u_{ij} = \left[\sum_{t=1}^{c} \left(\frac{\left\| x_j - v_i \right\|_2^2}{\left\| x_j - v_t \right\|_2^2} \right)^{\frac{1}{m-1}} \right]^{-1}, i = 1, 2, \cdots, c; j = i = 1, 2, \cdots, n \tag{9}$$

The basic steps of IFCM algorithm are shown in Figure 2:

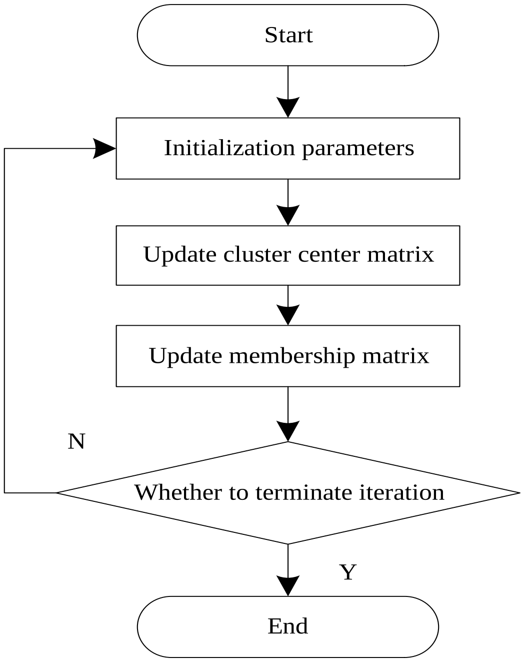

Start
Initialization parameters
Update cluster center matrix
Update membership matrix
N
Whether to terminate iteration
Y
End

Fig. 2. IFCM Algorithm Process.

1) Initialize algorithm parameters, namely set an iteration stopping threshold ε, a fuzzy clustering parameter m. The number of clustering is $c(2 \leq c \leq \sqrt{n})$ and the maximum iteration number is G. Initialize the membership degree matrix $U^{(0)}$.

2) Update a clustering center matrix: when that iteration time reach the $l(l = 1, 2, \cdots)$ time, update the clustering center matrix $U^{(l-1)}$ by using the clustering prototype calculation formulas (6) and (7) in combination with the membership degree partition matrix $\overline{V}^{(l)}$.

3) Update the membership degree matrix: update the partition membership degree matrix $U^{(l)}$ by using formulas (8) and (9).

4) Iteration termination condition of the algorithm: when the number of iterations reaches the maximum, or $\forall i$, k, $\max\left|u_{ik}^{(l)} - u_{ik}^{(l-1)}\right| < \varepsilon$, the iteration is stopped, then the IFCM algorithm stops and outputs the partition matrix U and the clustering prototype matrix $\overline{V}$; otherwise $l = l + 1$, return to step (2) [18].

C. The Autoencoder Algorithm Normalization Flow

The basic flow of the auto-encoder algorithm is as follows:

1) Input normalization. All the attributes of the input data is transformed into the number between [0, 1] to improve the flexibility of the model and eliminate the difference between the orders of magnitude of each attribute.

2) Initialize the auto-encoder model parameters. The parameters include the node number of input layer, the layer number of hidden layer, the node number of each layer of hidden layer, the encoder weight matrix, decoder weight matrix, center vector, maximum training times, and the model self-learning rate and parameters.

a) Input data. Sample points are drawn from the data set.

b) Random sampling in noisy data.

c) The SGVB gradient estimation.

d) Update parameters and weight.

e) Judge whether that parameter are converged.

f) Determination of termination. When the model is completely converged or reaches the maximum training times, the training is ended; otherwise, the step (3) is repeated.

IV. IVAEGAN-IFCM ESTABLISHMENT OF ALGORITHM

To solve the problem of incomplete data fuzzy clustering, an interval fuzzy clustering algorithm based on the IVAEGAN estimation (IVAEGAN-IFCM) is proposed to cluster the incomplete data sets. Through the IVAEGAN model, the incomplete data set is "restored" into a complete numerical data set, and the complete numerical data set is converted into an interval data set according to the interval rule proposed in this paper, and then the interval data set is analyzed by fuzzy clustering [19].

The specific algorithm flow of the IVAEGAN-IFCM algorithm is shown in Fig. 3.

1) Construct a nearest neighbor sample set for an incomplete data sample. The nearest neighbor samples are selected according to the nearest neighbor rule, and the nearest neighbor sample set of incomplete data is constructed.

2) Construct an attribute nearest neighbor sample interval of incomplete data. According to the interval rule proposed in this paper, the maximum and minimum values of missing attributes are determined in the nearest neighbor sample set of incomplete data to construct the nearest neighbor interval of attributes.

3) Determine the density of data samples in the neighborhood of incomplete data samples [20].

4) Input sample normalization. All data are converted into numbers between intervals$^{[0,1]}$, thus eliminating the difference of orders of magnitude between dimensions.

5) Initialize the IVAEGAN model. Initialize the network parameters in the IVAEGAN model, weight, bias value, maximum number of iterations, and training error.

6) Train the IVAEGAN model. The IVAEGAN model was trained on the complete data.

7) Fill in the missing attributes. The IVAEGAN model proposed in this paper estimates and predicts each missing data attribute in incomplete data, and at the same time obtains the estimation error of the IVAEGAN network for the complete attributes in the data set.

8) Interval data set: according to the interval type transformation rule proposed in this paper, all the data in the numerical data set are transformed into interval type, and then the interval type matrix is constructed.

9) Initialize the IFCM algorithm parameters. Initialize the membership degree matrix, and set the number of clustering categories, the number of cycles, the termination threshold and the fuzzy index.

10) Update the cluster center matrix. Update the clustering center matrix $V^{(l)}$ according to the clustering center matrix$U^{(l-1)}$.

11) Update that membership matrix. The statement $V^{(l)}$ updates the membership degree matrix$U^{(l)}$.

12) Algorithm condition judgment: when the number of iterations reaches the maximum, or $\max\left|U^{(l+1)} - U^{(l)}\right| \le \varepsilon$, the algorithm stops iterating; otherwise $l = l+1$, it returns to (10).

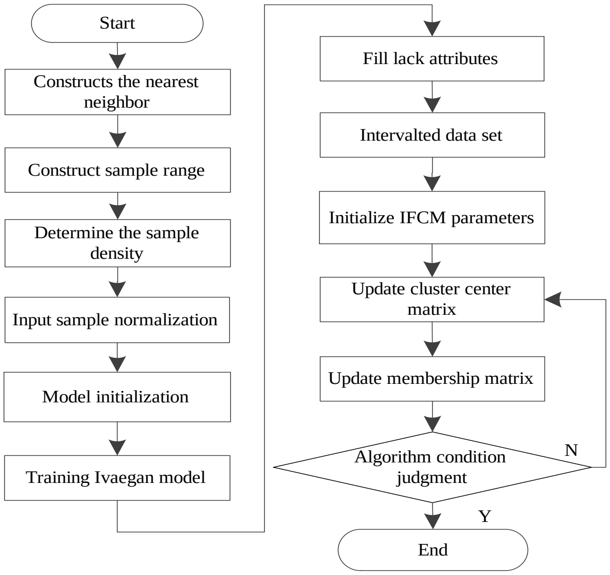

Start
Constructs the nearest neighbor
Construct sample range
Determine the sample density
Input sample normalization
Model initialization
Training Ivaegan model
Fill lack attributes
Intervalted data set
Initialize IFCM parameters
Update cluster center matrix
Update membership matrix
Algorithm condition judgment
N
Y
End

Fig. 3. IVAEGAN-IFCM Algorithm Process.

V. Experimental Analysis

A. Experimental Preparation

Aiming at the problem that the FCM clustering algorithm cannot directly use incomplete data for clustering analysis, this paper proposes a numerical incomplete data fuzzy clustering algorithm based on the IVAEGAN estimation. The missing attributes in the incomplete data are estimated and filled by the IVAEGAN model, and then the complete data set after recovery is analyzed by fuzzy clustering using the FCM clustering algorithm.

B. Construct the Artificial Dataset

In this experiment, the effectiveness of the algorithm is verified by artificial data sets, which are generated by the methods in the literature. Two artificial data sets are obtained. The artificial data set Test 1 contains two types of data. Each type has 1000 data samples, and the total number of samples is 2000 [21]. The artificial data set Test 2 contains three types of data, each type of data contains 800, 1000 and 2200 samples respectively, and the total number of data sets is 4000. Two artificial data sets subject to independent two-dimensional normal distribution are generated according to the above literature, and the conditional expectation and variance are as follows:

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

Where, the data generation parameters of manual data sets Test 1 and Test 2 are shown in Table II:

TABLE II. Artificial Datasets

	u_1	u_2	σ_1^2	σ_2^2
Test 1	3	3	2	2
	7	9	2	2
Test 2	28	24	6	6
	38	30	9	11
	50	42	17	20

According to the above generation method and parameter settings, generate a series of artificial data sets, in which the distribution of Test 1 is shown in Fig. 4, and the distribution of Test 2 is shown in Fig. 5.

In Fig. 4, the data in Test 1 is generated based on the first set of parameters. The two types of data are generated in equal quantities, and the data are evenly distributed, but the obvious differences of the data categories are maintained.

In Fig. 5, due to the different distribution and data amount of the three types of data, the first type of data is more concentrated, while the second and third types of data are more dispersed. The three types of data can still maintain relative independence, which is suitable for the validity verification of clustering algorithm.

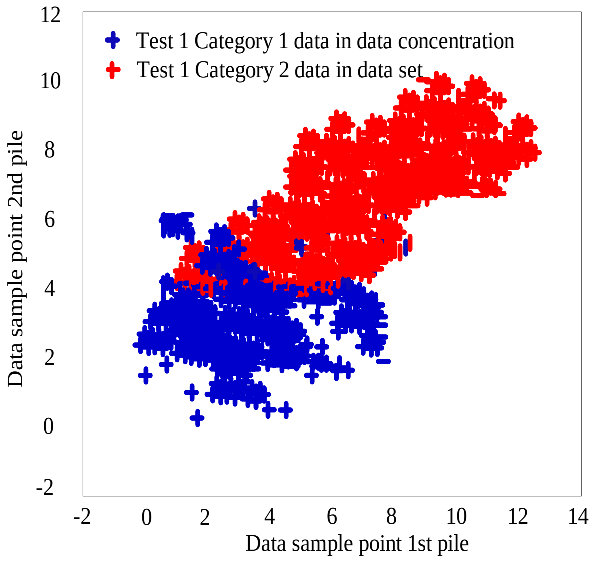


Fig. 4. Artificial Dataset Test 1.

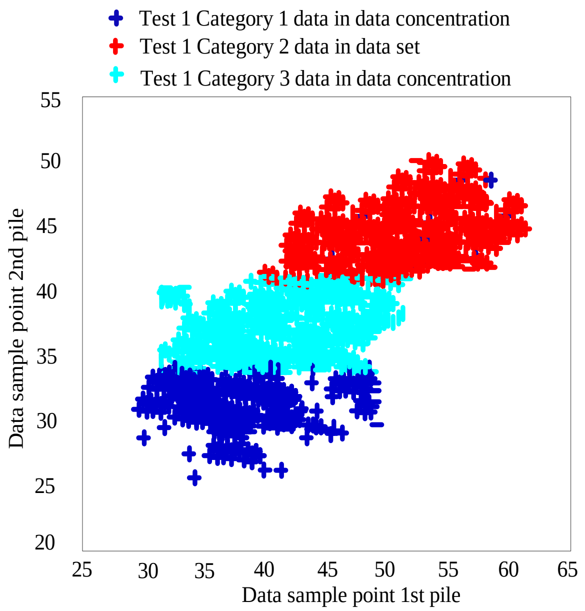


Fig. 5. Artificial Dataset Test 2.

C. *Incomplete Data Generation Rules*

In the whole data processing process, in order to make the incomplete data generated in the experiment closer to reality, the incomplete data set is generated by randomly discarding the data and randomly losing the complete data in different proportions set manually. In an incomplete data set, the missing attribute of the data sample is denoted by "?" to represent. The rules for generating data missing attributes of the incomplete data sets are as follows:

1) In an incomplete data set, the attribute values of a sample data cannot be completely lost, that is, if the data set is n-dimensional, the incomplete data in it can lose at most attributes, and at least one attribute of an incomplete data must exist.

2) In the incomplete data set, any one-dimensional attribute has at least one complete attribute value, that is, the attribute column of the data set cannot be empty to ensure the reliability of the valuation.

In the algorithm comparison experiment of this paper, the complete data strategy fuzzy C-means clustering algorithm (whole data strategy fuzzy C-means clustering) WDS-FCM) and the local distance strategy fuzzy C-means clustering algorithm (partial distance strategy fuzzy C-means clustering, PDS-FCM) are rejection methods [22].

The remaining two comparison algorithms optimize the complete strategy fuzzy c-means clustering algorithm (optimal completion strategy fuzzy C-means clustering, OCS-FCM) and the nearest prototype strategy fuzzy C-means clustering algorithm (Nearest Prototype Strategy fuzzy C-means clustering, NPS-FCM) belong to the valuation method [23].

The IVAEGAN-IFCM algorithm proposed in this paper uses the IVAEGAN model to estimate the incomplete missing attributes. The IVAEGAN model performs feature extraction and data generation on the incomplete data set. Meanwhile, the model combines the median of the neighbor data as the conditional label to make the estimation range more accurate and efficient. After the IVAEGAN model performs estimation filling on the incomplete data set, the obtained complete data set completes the fuzzy clustering, which improves the clustering accuracy [24]. Iris, Bupa and Breast data sets are used in the experiment. Each data set contains 500 samples, 11 feature categories and 9 attributes.

According to the average number of iterations in Tables III to V, the algorithm in this paper can converge stably after multiple iterations under different missing rates of each data set. Compared with other clustering algorithms, the algorithm does not achieve the best results, but it can still achieve stable results after a certain number of iterations.

TABLE III. AVERAGE ITERATION TIMES OF IRIS

Missing rate /%	WDS-FCM	PDS-FCM	OCS-FCM	NPS-FCM	Algorithm
5	25.1	25.3	27.5	27.8	30.5
10	25.9	26.3	27.9	28.2	30.8
15	26.1	26.7	28.6	29.3	31.4
20	26.8	27.1	28.8	29.7	31.8

TABLE IV. AVERAGE NUMBER OF ITERATIONS OF BUPA

Missing rate /%	WDS-FCM	PDS-FCM	OCS-FCM	NPS-FCM	Algorithm
5	33.7	36.2	36.6	36.7	36.5
10	35.6	37.4	39.1	37.7	39.3
15	37.3	35.9	38.6	38.9	42.4
20	38.6	36.5	36.8	40.3	46.2

TABLE V. AVERAGE ITERATION TIMES OF BREAST

Missing rate /%	WDS-FCM	PDS-FCM	OCS-FCM	NPS-FCM	Algorithm
5	15.3	15.2	16.4	15.6	18.4
10	15.9	16.3	16.8	15.7	18.9
15	16.1	16.8	17.2	16.3	22.3
20	16.9	17.4	18.6	17.2	24.6

According to the experimental results, the algorithm in this paper is relatively better, in the case of small data loss, such as 15% loss, the clustering effect of the numerical estimation method is better, which is obviously better than other algorithms.

From the standard deviation of clustering error scores in Tables VI to VIII, the algorithm in this paper can maintain a low standard deviation of clustering error scores under different missing rates of different data sets, which reflects the stability of the algorithm. In some cases, such as the case of less missing rate, the optimal solution cannot be obtained, and in other cases, the best results can be obtained.

TABLE VI. STANDARD DEVIATION OF IRIS CLUSTERING ERROR SCORE

Missing rate /%	WDS-FCM	PDS-FCM	OCS-FCM	NPS-FCM	Algorithm
5	1.24	1.56	2.02	1.86	2.03
10	2.15	1.75	1.95	1.78	1.78
15	2.08	1.89	2.04	169	1.93
20	2.09	1.92	2.01	2.07	2.04

TABLE VII. STANDARD DEVIATION OF BUPA'S CLUSTERING ERROR SCORE

Missing rate /%	WDS-FCM	PDS-FCM	OCS-FCM	NPS-FCM	Algorithm
5	2.03	2.36	2.12	2.36	2.04
10	1.83	1.89	2.05	2.04	2.12
15	2.51	1.38	1.89	2.01	2.23
20	2.18	1.94	1.95	1.98	2.64

TABLE VIII. STANDARD DEVIATION OF BREAST CLUSTERING ERROR SCORE

Missing rate /%	WDS-FCM	PDS-FCM	OCS-FCM	NPS-FCM	Algorithm
5	3.54	3.12	3.45	2.89	4.31
10	3.57	2.92	3.23	2.96	4.56
15	3.26	3.02	3.21	3.25	4.89
20	3.58	3.08	3.04	3.12	4.97

Test results show that the algorithm in this paper is better than other algorithms on the whole. Compared with other algorithms, the performance of our algorithm is improved by 4%. 3% when the missing rate of the data set is low, such as the missing rate of 5% and 10% of the small data loss. With the increase of the missing rate, the test performance of the sample is also gradually improved. It is proved that the algorithm proposed in this paper can fully show the uncertainty of incomplete data estimation by using interval data, and has superior performance in the process of missing data filling and clustering analysis.

The WDS-FCM algorithm eliminates all the incomplete sample data in the data set and only performs fuzzy clustering on the complete sample data. The clustering accuracy of the algorithm will be greatly affected when the proportion of incomplete samples increases. The PDS-FCM algorithm replaces the Euclidean distance in FCM with the local distance formula, and only adds the iterative operation to the complete attributes in the processing of incomplete sample data, without considering the missing attribute information of incomplete samples, which does not give full play to the information value of incomplete samples. Therefore, both the WDS-FCM and the PDS-FCM do not make full use of the effective information value of the mining incomplete data. The IVAEGAN-FCM algorithm proposed in this paper does not delete and abandon incomplete data, but reconstructs the model to fill in the missing data, so that the data set can be "restored" to a complete data set. The training samples are used to train the IVAEGAN model, and the missing attributes of each incomplete attribute are estimated and filled to obtain the "recovered" complete data set, and then the complete data set is subjected to fuzzy clustering.

VI. CONCLUSION

In this paper, according to the incomplete data formed in the power system, a numerical incomplete data fuzzy clustering algorithm based on the improved IVAEGAN estimation is proposed to solve the problem that the traditional clustering algorithm cannot directly use the incomplete data. A new fusion model is constructed by combining VAE and GAN to estimate and fill the incomplete data. The main work includes:

1) Fill the incomplete data attributes to realize the estimation clustering. A fuzzy clustering algorithm for numerical incomplete data based on the improved IVAEGAN estimation is proposed.

2) Construct a nearest neighbor sample set for that incomplete data according to the nearest neighbor rule, generating a model VAE and a model GAN, and constructing an IVAEGAN model.

3) Construct a missing attribute label by using that median value of the attribute of the nearest neighbor sample set so that the IVAEGAN model can obtain more effective information, and the estimation accuracy is improved.

4) UCI data sets and two artificial data sets are used for comparative experiments to verify the effectiveness of the algorithm, and the effectiveness of the algorithm is summarized in depth.

The model in this paper uses the gradient integral to update the parameter values, which is fast and has high computational complexity, and has obvious effect on improving the training speed. Compared with other models, this paper uses the attributes of the nearest neighbor sample set to construct condition variables, and other construction methods of condition variables are more effective, which further optimizes the space and improves the theoretical and practical basis of data integrity research. The research results of this paper are very important to fully mine the effective information in the incomplete data of power system, and play an important role in ensuring the normal operation of power system.

In the future work, more optimization methods will be used to optimize the parameter update, such as the wolf pack optimization algorithm. The GAN model is prone to the phenomenon of gradient explosion, center collapse and training non-convergence to further enhance the space for improvement.

REFERENCES

[1] Zhong G Q, Gao W, Liu Y B, Yang Y Z. Generative adversarial networks with decoder–encoder output noises[J]. Neural Networks,2020,127:19-28.

[2] Yoon J ,Jordon J, Mihaela S. GAIN: Missing Data Imputation using Generative Adversarial Nets[J]. arXiv preprint arXiv:1808.02920, 2018.

[3] Martin A, Soumith C, Bottou L. Wasserstein GAN[J]. arXiv preprint arXiv:1701.07875v3, 2018.

[4] Mihaela R, Balaji L, David W, Shakir M. Variational Approaches for Auto-Encoding Generative Adversarial Networks[J]. arXiv preprint arXiv:1706.04987v2,2017.

[5] Du C D, Li J P, Huang L J, He H G. Brain Encoding and Decoding in fMRI with Bidirectional Deep Generative Models[J]. Engineering,2019,5(5):948-953.

[6] Chen S M, Yu J B, Wang S J. One-dimensional convolutional auto-encoder-based feature learning for fault diagnosis of multivariate processes[J]. Journal of Process Control, 2020, 87:54-67.

[7] Ahmad M K, Mehmet S G, Mehmet R T, Hilal K. A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing [J]. Biocybernetics and Biomedical Engineering,2019,39(1):148-159.

[8] Tang L J, Zheng S C, Zhou Z G. Estimation and inference of combining quantile and least-square regressions with missing data [J]. Journal of the Korean Statistical Society, 2018, 47(1):77-89.

[9] Liu X J, Zhang H, Niu Y G, Zeng D L. Modeling of an ultra-supercritical boiler-turbinesystem with stacked denoising auto-encoder and long short-term memory network [J].Information Sciences, 2020, 525(1):143-152.

[10] Chen G, Wang H, Jian T, Xu C, Sun S. Method for denoising and reconstructing radar HRRP using modified sparse auto-encoder[J]. Chinese Journal of Aeronautics, 2020,33(3):1026-1036.

[11] Gao Z S, Shen C, Xie C Z. Stacked convolutional auto-encoders for single space target image blind deconvolution[J]. Neurocomputing,2018,313(3):295-305.

[12] Liang Y, Ke S, Zhang J, et al. Geoman: Multi-level attention networks for geo-sensory time series prediction[A]. International Joint Conference on Artificial Intelligence[C]. 2018: 3428-3434.

[13] Han M, Zhong K, Qiu T, et al. Interval type-2 fuzzy neural networks for chaotic time series prediction: a concise overview[J]. IEEE Transactions on Cybernetics, 2018, 49(7):2720-2731.

[14] Wang H, Yang Z, Yu Q, et al. Online reliability time series prediction via convolutional neural network and long short term memory for service-oriented systems[J]. Knowledge-Based Systems, 2018, 159:132-147.

[15] Araújo R A, Nedjah N, Oliveira A L I, et al. A deep increasing–decreasing-linear neural network for financial time series prediction[J]. Neurocomputing, 2019, 347:59-81.

[16] Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: A strong baseline[A]. International Joint Conference on Neural Networks[C]. 2017:1578-1585.

[17] Ma Q, Zhuang W, Shen L, et al. Time series classification with Echo Memory Networks[J]. Neural networks, 2019, 117:225-239.

[18] Fawaz H I, Forestier G, Weber J, et al. Deep learning for time series classification: a review[J]. Data Mining and Knowledge Discovery, 2019, 33(4):917-963.

[19] Ma Q, Zheng J, Li S, et al. Learning Representations for Time Series Clustering[A]. Neural Information Processing Systems[C]. 2019:3776-3786.

[20] Shen L, Ma Q, Li S. End-to-end time series imputation via residual short paths[A]. Asian Conference on Machine Learning[C]. 2018:248-263.

[21] Cao W, Wang D, Li J, et al. Brits: Bidirectional recurrent imputation for time series[A]. Neural Information Processing Systems[C]. 2018:6775-6785.

[22] Luo Y, Cai X, Zhang Y, et al. Multivariate time series imputation with generative adversarial networks[A]. Neural Information Processing Systems[C]. 2018: 1596-1607.

[23] Luo Y, Zhang Y, Cai X, et al. E 2 GAN: end-to-end generative adversarial network for multivariate time series imputation[A]. International Joint Conference on Artificial Intelligence[C]. 2019:3094-3100.

[24] Zhang J, Yin P. Multivariate Time Series Missing Data Imputation Using Recurrent Denoising Autoencoder[A]. International Conference on Bioinformatics and Biomedicine[C]. 2019:760-764.

MCBRank Method to Improve Software Requirements Prioritization

An Empirical Investigation

Sabrina Ahmad[1], Riftika Rizawanti[2], Terry Woodings[3], Intan Ermahani A. Jalil[4]
Fakulti Teknologi Maklumat Dan Komunikasi, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia[1, 4]
Fakultas Teknologi Informasi dan Komunikasi, Universitas Teknologi Mataram, Nusa Tenggara Barat, Indonesia[2]
School of Physics, Math and Computing, The University of Western Australia, Perth, Australia[3]

***Abstract*—Software Requirements Prioritization is an important issue that has a more profound effect on the overall quality of software development. Application of software requirements prioritization provides benefits to minimize risks in software development so that the most important and most impactful requirements are given priority. This paper presents a proposed software requirements prioritization method named MCBRank, which incorporates renowned MoSCoW Method and Case-Based Ranking to improve prioritization correctness. It elaborates on the implementation of MCBRank in an empirical investigation to determine software requirements prioritization for a potential e-library system. The investigation allows the software requirements prioritization process to be implemented by using the proposed MCBRank method. A role-playing empirical investigation with 30 respondents prioritized 31 software requirements, and the results were measured by Cohen Kappa. The kappa results show that MCBRank achieves a better agreement towards the Gold Standard with kappa value of 0.60. Therefore, the investigation results support that MCBRank improves the importance of ranking correctness, representing the stakeholders' wants and the organization's actual needs for the potential e-library system.**

***Keywords*—Requirements prioritization; requirements engineering; software engineering; empirical software engineering**

I. Introduction

Requirements play an essential part in software development as they lay a foundation for a system to be developed [1]. Among many factors influencing the success or failure of software development, prioritization seems vital as it is responsible for determining the value of various requirements proposed by many stakeholders based on specific criteria defined from their usefulness for the final software product [2]. Thus, requirements prioritization is an essential part of requirements management. It is responsible for identifying the subset from many candidate requirements to maximize the fulfilment of various limitations such as resources availability, constrained timeline, and budget.

Thus, the main objective of the prioritization is to ensure that the customers get the software system that satisfies their needs with utmost value within limited resources. This is supported by a study [3] that stated Software Requirements Prioritization (SRP) could affect several factors in software development such as value, time to market, cost, and, most importantly, user satisfaction. Using SRP, the requirements will be prioritized based on human choices and by analyzing several factors that can be the reason for a product to fail or succeed.

The importance of having the correct prioritized requirements which fulfil the stakeholders' needs [4] and the difficulty of getting a good set of requirements that satisfy the business value motivate this research to propose an enhanced MoSCoW prioritization method called MCBRank. The proposed method incorporates renowned MoSCoW method and Case-Based Ranking method to improve the prioritization correctness. In this paper, an empirical method is adapted to evaluate the MCBRank effectiveness to prioritize software requirements. Therefore, this paper presents the empirical investigation of the MCBRank prioritization method based on a case study. The empirical result showed that the MCBRank method improves the software requirements prioritization correctness which represent the stakeholders' wants and the organization's actual needs for the potential e-library system.

Following the introduction, Section II presents the literature review. Next, Section III presents the proposed prioritization method, and Section IV explains the empirical investigation methodology. This is followed by Section V that discusses the empirical investigation results. Finally, Section VI concludes the paper.

II. Literature Review

Requirements prioritization refers to the process of choosing the correct set of specifications from a mass of overlapping and contradictory demands from various stakeholders involved in a software development project. Prioritization is one of the critical steps in system development to make meaningful decisions to scrutinize essential requirements for realization. This is supported by a study which stated that requirements prioritization has an enduring partnership with several other important technical practices such as interaction with the requirements analysis, requirement engineering, and general software engineering practices [5].

Gilb and Maier [6] stated that "priority is the relative right of a requirement to the usage of restricted (or scarce) capital." In this description, "capital" includes all kinds of capital, including time, money, and human capital. In other terms, we can have anything of the infinite system; then, there is no need to consider the requirements. However, ventures usually face finite capital such as short deadlines, restricted budgets, limited human capacity, and fixed technologies. Consequently, proposals sometimes include more requests than what can be executed in one release period. Thus, requirements prioritization enables the project planners to choose the final requirements to be implemented under resource constraints.

Various requirements prioritization techniques exist, but using the same technique to assign requirements for multiple stages may result in restricted benefits [7]. Different requirements prioritization techniques involve various properties, and therefore selecting the most optimal technique will optimize the gains gained at specific points [8]. These techniques enable the software development team to prioritize software requirements with high priority [9]. According to a study [5], there are many methods used to classify requirements. These prioritization methods may be classified as nominal scale, ordinal scale, ratio scale, machine learning, and hybridization, as shown in Fig. 1.

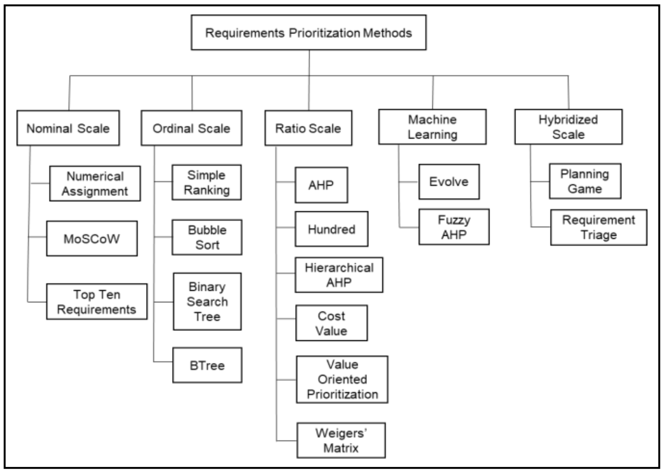


Fig. 1. Requirements Prioritization Method [19].

Other than methods, the elements being considered during the prioritization process are also essential. Hujainah et al. [10] provide a comprehensive investigation of the interrelated elements that need to be considered to formulate practical requirements prioritization technique. The findings reveal four interrelated elements that should be considered in developing RP techniques to secure an effective prioritization process. These elements are criteria for requirements, stakeholders, procedures, and implementation, which play important roles in prioritizing requirements. To incorporate the interrelated elements for prioritization, further research [11] proposed a semi-automated scalable prioritization technique to improve the prioritization efficiency for a large-scale project.

The fact that stakeholders’ preferences need to be considered is also supported by Yaseen et al. [12]. It is noted that large numbers of requirements are likely to be developed based on the customers' preferences stated in the early stage of the requirements engineering phase. However, diverse stakeholders need to be considered since specific requirements are important for particular customers but not others [13]. Furthermore, customers may generally understand what they want but do not have a specific picture of precisely what is needed for a well-functioning software system. Therefore, requirements prioritization may assist the development team in shortlisting the requirements because essential requirements should be presented to the stakeholders as quickly as possible. In addition, should conflicting requirements surface, requirements prioritization can be performed to resolve the conflicts. This is also supported by Ma [7], who stated that requirement prioritization refers to the process of choosing the correct set of specifications from a multitude of overlapping and contradictory demands from various stakeholders involved in a software development project.

Achimugu et al. [14] performed a systematic review of 48 Software Requirements Prioritization methods and found that MoSCoW is the most cited and utilized prioritization method. Miranda [15] noted that the MoSCoW is a more straightforward method of obtaining information from customers, meaning that customers better understand what is being asked of them and thus provide the development team with more meaningful and valuable information. Moreover, the MoSCoW is suitable for iterative development such as "agile software development" [16, 17, 18], which allows collaborative requirements prioritization between stakeholders [19]. This collaborative effort provides the customers with a product of a maximized business value [20]. On the other hand, the MoSCoW needs more time as detailed information is required to provide a relative value for each requirement. Besides MoSCoW technique is a numerical assignment technique that needs more effort to solve conflicts between analysis and stakeholders' viewpoints [21].

Meanwhile, Avesani et al. [22] proposed a framework called Case-Based Ranking that can reduce the acquisition effort by combining human reference elicitation and automatic preference approximation. The result shows improvement in requirement prioritization accuracy and a trade-off between experts’ elicitation efforts. During the requirements analysis process on deciding which requirements to develop, different methods are used to select the correct requirements due to the system development team's preferences and work nature.

While MoSCoW and CBRank have received a great deal of attention in the literature, MCBRank method is designed to emphasize the strength of both techniques to improve software requirements prioritization.

III. The Proposed Method

According to Yaseen et al. [23], it is necessary to recognize requirements' importance and priority to assist the developers in expediting the system development process. MCBRank enables two-layer prioritization to improve the correctness in terms of importance ranking. In this research design, the importance is represented by the key stakeholders’ cumulative needs for the best of the system to be developed. Fig. 2 gives an overview of the MCBRank. Firstly, all candidate requirements which come from multiple stakeholders are listed. Then the key stakeholders are required to classify each requirement based on the adapted MoSCoW method on five points scale. The MoSCoW classification of M (Must have), S (Should have), C (Could have), and W (Would have) are assigned with numbers as listed below. ‘Must not have this’ is added into the scale to allow stakeholders to indicate the requirements they do not want to be realized.

5 - Must have this.

4 - Should have this if at all possible.

3 - Could have this if it does not affect anything else.

2 -Will not have this time but would like in the future.

1 - Must not have this.

Following that, the classification score based on the scale will be used to classify all the requirements into M (Must have), S (Should have), C (Could have), and W (Would have). At this point, the unwanted requirements will be discarded.

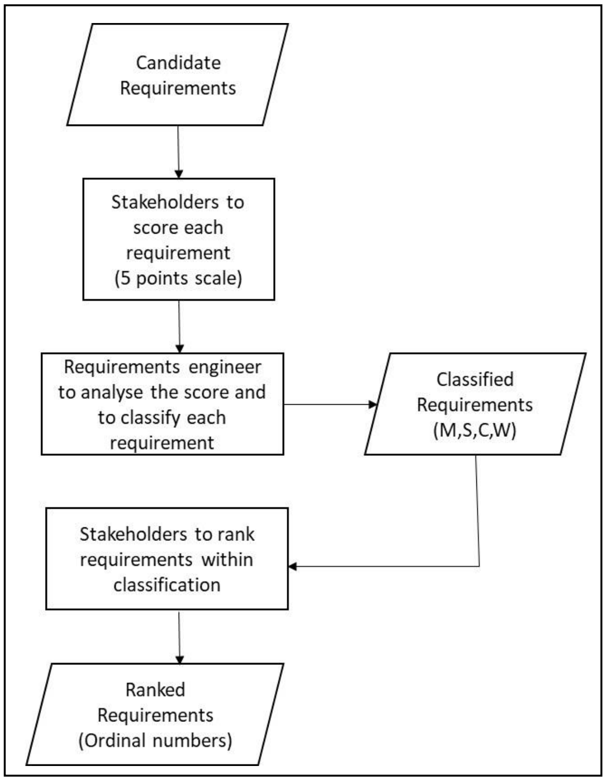

Candidate Requirements
Stakeholders to score each requirement (5 points scale)
Requirements engineer to analyse the score and to classify each requirement
Classified Requirements (M,S,C,W)
Stakeholders to rank requirements within classification
Ranked Requirements (Ordinal numbers)

Fig. 2. Overview of MCBRank.

Next, within the classification, each requirement will be ranked using ordinal numbers. The majority rank will be the position of the requirement. If the majority rank is the same for two or more requirements, a smaller accumulative score given by the participating stakeholders will be in a higher rank. Finally, a new ranked requirements list is produced.

This ranking allows stakeholders to prioritize the requirements, delineate and narrow the scope of work to acquire focus. Priorities make it possible to measure how important the stakeholders feel about each requirement concerning a software solution to meet their needs. The prioritized requirements successfully narrow the focus, which helps in group agreement. Through priority [12], if it is impossible to develop all project requirements, it is feasible to discriminate the most critical requirements to the stakeholders. This means that a project that does not meet all of its requirements can still be of high value if it meets the stakeholders' most important requirements.

A. *An Example*

Assume that there are ten requirements proposed by three key stakeholders numbered as Req. 1 until Req. 10.

Step 1: The key stakeholders to score the listed requirements based on importance (5 points scale).

Requirements	**Stakeholder 1**	**Stakeholder 2**	**Stakeholder 3**
Req 1	5	5	4
Req 2	4	4	5
Req 3	5	4	5
Req 4	5	3	5
Req 5	3	3	4
Req 6	4	5	5
Req 7	5	2	1
Req 8	3	4	3
Req 9	2	3	3
Req 10	1	2	1

Step 2: The requirements engineer to analyze the score and to classify each requirement. In order to do this, the requirements engineer need to analyze the requirements and take into consideration the technical knowledge on top of just importance classified by the stakeholders. This is required to ensure the essential services for the said system is not neglected.

Req.	**Stakeholder 1**	**Stakeholder 2**	**Stakeholder 3**	**Group**
Req 1	5	5	4	M
Req 2	4	4	5	S
Req 3	5	4	5	M
Req 4	5	3	5	M
Req 5	3	4	4	S
Req 6	4	5	5	M
Req 7	5	2	1	W
Req 8	3	4	3	C
Req 9	2	3	3	C
Req 10	1	2	1	discard

Step 3: Within each classification group, the key stakeholders to rank requirements based on ordinal numbers. The majority will be the rank of the requirements. Note that, the ordinal number will follow through based on classification group priority. For example, Req 2 is ranked as 1 within Group S, then the number must be after the last number in the group M. Therefore, Req 2 is ranked as 5.

Group	**Req**	**S 1**	**S 2**	**S 3**	**Rank**
M	Req 1	2	1	2	2
	Req 3	1	2	1	1
	Req 4	3	4	3	3
	Req 6	4	3	4	4
S	Req 2	1	2	1	5
	Req 5	2	1	2	6
C	Req 8	2	2	1	8
	Req 9	1	1	2	7
W	Req 7	1	1	1	9
-	Req 10	-	-	-	-

IV. Empirical Investigation methodology

An empirical investigation method was carried out to evaluate the effectiveness of the proposed MCBRank Method. It is a role-playing investigation to test if the requirements prioritization is improved following the implementation of MCBRank.

A. The Underlying Concept

The underlying concept includes an explanation and rationale for the terms and approaches applied.

- Stakeholders are terms that refer to any person or group directly or indirectly affected by the system. Stakeholders include end-users who interact with the system and everyone in the organization who may be affected by its installation. Other system stakeholders may be engineers who develop or maintain related systems, business managers, domain experts, and union representatives. These stakeholders are the key people representing the interests of their group. They may include end-users, system owners, and managers who work together and are actively involved in decision-making to reach mutually satisfactory agreements. However, it is neither appropriate nor possible to have all system stakeholders in acquiring requirements. Therefore, during the role-play investigation, this project will involve key stakeholders to represent users and administrators.

- The Gold Standard (GS) is a term used to describe the theoretical idea of the best requirements set that can exist for a system. In this requirements prioritization study, the 'ideal' set represents the right ranked requirements based on importance, representing the cumulative key stakeholders' needs and the actual need of the organization. This is because a sound system should fulfill what the stakeholders want and what the customers describe is what they want and what the organization needs from the system to be developed. However, there is no way to test this benefit in the short term. That is why this study presents the GS to represent a set of right ranked requirements based on importance from the domain experts' point of view. The domain experts are familiar with the crucial need of the system to fit the system's purposes of existence and understand the system's business value. The GS is determined by the candidate requirements and the collaborative efforts of researchers and several domain experts in the field. The experts are library practitioners ranging from thirty-five to fifty years old, with more than ten years of experience working in the library for Universiti Teknikal Malaysia Melaka. It is established as a benchmark for the best possible requirements with the correct importance rank for the particular case study; e-library system. The results from the investigations that will be carried out will be measured against the GS.

B. Gold Standard Formulation

This subsection describes how the researcher obtained the Gold Standard (GS). The GS is an ideal set of ranked requirements based on importance from the point of view of domain experts. The GS is obtained through an expert judgment technique that involves a group of domain experts. GS is a term used to describe the theoretical idea of the ideal set of well-ranked requirements based on importance for a system. GS was developed carefully by identifying some requirements that contain all the primary requirements required for the system. Researchers are aware that it is impossible to find a precise solution, but estimates can be determined based on the analysis of several experts.

Subjects involved in determining GS are experts who are library practitioners with criteria aged 30-50 years, have more than ten years of experience in related fields, and come from the Universiti Teknikal Malaysia Melaka. The procedures carried out are:

- In the first step, the researcher prepares a set of instruments containing an introduction, instructions, e-library scenario description, and a decision form that contains 31 candidate requirements to collect expert opinions.

- In the second step, screening and determining experts based on predetermined demographic components were conducted. This study involved six experts in determining GS.

- In the third step, the experts were given three days to examine the GS determination instrument to get familiar with the introduction, instructions, descriptive scenario, and a list of candidate requirements. Experts will get an overview of the background system from which the requirements were obtained through descriptive scenarios. Then the experts were asked to provide a rating with a Likert scale of 1-5 on the decision form. The scale of the assessment technique used in this study was adapted from the MoSCoW method.

Following the three steps above, the researcher gets the results of the requirements that have been ranked according to the highest to lowest rated. These properly ranked requirements are determined as GS. The value of each requirement is obtained from the total rate given by the six experts. Sorting has no problem when a requirement has different total values and does not get an equal value. For the first and second ranks, there were no obstacles. As for rank three and four, the total value for requirements R6 and R29 are the same, but since R29 has a higher number of five-point rating which means that more people rate R29 as 'Must have, R29 is placed in the third rank followed by R6 in the fourth rank. This also applied to the requirements R7, R5, and R22; R9 and R27; R17 and R23. As for requirements R26 and R12, which have the same total value and the same number of five-point ratings, R26 was decided to be in a lower rank than R12 because there exists a lower point rating for R26. Fundamentally, for the rest of the requirements, if the total value is the same and if the five-point rating is the same, then a lower point rating, if it exists for one of the requirements, will determine the ranking of the said requirements.

The formation of the GS in this investigation was done carefully since the GS represents an ideal set of requirements with the right importance ranking (theoretically represents both the stakeholders' wants and the organization's needs). The right rank of importance will determine the most crucial requirements to be realized to reduce the timeline or budget shortage. Thus, the output of the MCBRank will be measured against the GS to evaluate its effectiveness.

C. Investigation Method

1) Investigation materials: The candidate requirements are derived from the e-Library descriptive scenario. Ideally, the candidate requirements are proposed by independent stakeholders (or by a group of stakeholders) [24], represented by participants who play roles as a system's stakeholders in this investigation. However, the investigation focuses on the requirements prioritization effort with the proposed method and the researchers' candidate requirements prepared in advance. Requirements are built on the experience of researchers in using the system. They are improved by academics who previously have years of experience in developing systems and working on requirements. For prioritization purposes, candidate requirements are constructed to represent the interests and needs of various stakeholders. Candidate requirements include essential system features and additional features with a mix of quality values. In addition, the requirements also meet SMART concept; (S)pecific, (M)easurable, (A)greed, (R)ealistic, and (T)ime-bonded [3]. The number of candidate requirements is balanced and adjusted to the interests of specific key stakeholders. This is generally based on the purpose of the investigation. The number of candidate requirements was set for the experiment considering its time to rank effectively.

2) Population and sample: The investigation was carried out with 30 people, and the role-playing technique was exercised. Each was given a role as a user (students, lecturers, university staff) and administrators (librarians). Participants who carried out the investigation had a background in software engineering knowledge and were between 20 to 30 years old. Several of them have work experience in software development. To avoid bias, the roles they played in the investigation were randomized. Also, role-playing investigations always attend to whether the participants are actually playing a role or are merely including their judgments. It is expected that each participant will be more committed to individual priorities. Participants are given an exact role to minimize that possibility and given instructions and guidelines on playing system stakeholders' positions. Since participants may also have limited knowledge to precisely rank all candidate requirements concerning their impact on the acceptance of the software system [19], scenario descriptions and candidate requirements have been given to them in advance. This reading material is helpful because the scenario description describes the system needs and concerns of different stakeholders. In addition, the candidate requirements are carefully tailored to the needs of specific stakeholders.

3) The investigation protocol: For investigation, participants were divided into two groups. Each group consisted of thirty participants to rank candidate requirements based on the role they were playing. One group was a control group to exercise the MoSCoW prioritization method, and the other group exercised the MCBRank prioritization method. The results of the investigation are a prioritized list of software requirements.

Initially, during the investigation process, all participants were given scenario descriptions of potential e-libraries. Furthermore, a briefing was conducted on the experimental background knowledge through online communication. This is followed by instructions supported by an overview of examples of step-by-step activities. The assignment of participants is tailored to the role that will be played during the investigation. The participants were given a day to understand the roles and descriptions of the scenario prepared for them. The next day, participants were given a google form link and asked to prioritize the candidate requirements based on the MoSCoW method and MCBRank method. The participants' prioritization effort was then collected and analyzed. When all the participants had completed the assignment, a feedback form was given to them to learn about the strengths and weaknesses of the investigation for future references.

D. Justification of Results Measures

A study [25] introduced the concept of a perfect set of requirements (here termed the 'Gold Standard'), and measurements of progress towards such an ideal are made. Cohen's Kappa [26] is used to measure agreement between two assessors. Cohen's Kappa is an inter-rater reliability index commonly used to measure the level of agreement between two sets of dichotomous ratings or scores. It is generally a more robust measure than simple percent agreement calculation since it considers the agreement occurring by chance. Here, Cohen's Kappa is used to measure the agreement between requirements prioritized using MCBRank and the Gold Standard. The two raters are the group exercising the MCBRank and the domain experts who provide the Gold Standard. Kappa values show the movement of the agreement towards the Gold Standard. The nearer the agreement to the Gold Standard, the better quality it is. Therefore, the suitability of this metric is discussed, and its relevance to this study is stated.

Each participant prioritizes each requirement by using the MoSCoW or MCBRank methods. The prioritized requirements resulting from both methods were evaluated against the GS using the Cohen Kappa. Cohen's Kappa is used to measure the distance or degree of agreement between the GS and a set of requirements obtained through the application of the MCBRank in the investigation. If the agreement between results from the MCBRank and GS is better than the agreement between MoSCoW and GS, then the MCBRank method has succeeded in improving the level of correctness in requirements prioritization.

V. Results and Discussion

The empirical investigation evaluates whether the enhanced MoSCoW method named MCBRank can improve software requirements prioritization in terms of the correct rank of importance. Role-playing technique as key stakeholders to the e-library system was exercised in this study where the participants were to prioritize the candidate requirements using the proposed method. The purpose of this study is to evaluate the MCBRank method.

A control group was formed to exercise the MoSCoW method to be compared with the proposed method (MCBRank). The results obtained from the MoSCoW method and MCBRank method were both calculated. The level of agreement of both methods with GS was calculated using the Cohen Kappa formula.

The formula to calculate Cohen's kappa for two raters is:

$$k = \frac{P_{o-}\, p_e}{1 - \ p_e}$$

where:

P_o = total number of requirements that have the same rank position in GS.

p_e = total number of expected similarities that would occur if the observers were statistically independent.

Cohen's kappa measures the agreement between two raters who classify several items into several mutually exclusive categories. Landis and Koch [27] mention that labels with appropriate kappa ranges should be used to maintain consistent terminology when describing the relative agreement strength associated with kappa statistics. Table I presents the agreement strength based on Cohen Kappa.

The prioritized requirements obtained from the MoSCoW method were calculated, and the agreement to the GS is presented in Table II.

Based on the guideline adapted from Landis & Koch [27], a kappa (κ) of .067 represents a poor strength of agreement. Furthermore, since p = .042 (which means p > .0005), our kappa (κ) coefficient is not statistically significant. Meanwhile, Table III presents the measure of agreement between prioritization obtained from the MCBRank method and GS.

Meanwhile, a kappa (κ) of .600 represents a moderate strength of agreement. Furthermore, since p = .000 (which means p < .0005), our kappa (κ) coefficient is statistically significant.

Kappa values show the movement of the agreement towards the Gold Standard. The nearer the agreement to the Gold Standard, the better quality it is. The investigation was carried out to determine whether there was an agreement between the prioritized requirements achieved through the MoSCoW method to the GS and those achieved through the MCBRank to the GS. There is a poor agreement between the MoSCoW method and the GS, κ = .067, p> .0005. On the other hand, the kappa value shows a moderate strength agreement between the MCBRank method and the GS, κ = .600, p <.0005. Through the investigation, this research demonstrates that the set of prioritized requirements obtained using the MCBRank method moves closer towards the ideal Gold Standard compared to the prioritized requirements obtained using the MoSCoW method. Therefore, requirements prioritization in terms of the correct ranked requirements based on importance is better achieved through MCBRank.

A. Limitations

The research was conducted in a limited time frame, and therefore there are several limitations in the investigation conducted as stated below:

- It is ideal to have candidate requirements proposed by the participants. However, due to limited execution time and the need to focus on the prioritization exercise, the researchers carefully prepared the candidate requirements based on a specific case study. Ample time is given for the participants to understand the requirements before the prioritization activity.
- It is best to have a complete set requirement for at least an industrial small size system to investigate the requirements prioritization performance [28]. However, the candidate requirements were reduced to allow sufficient time to exercise prioritization activity. The researchers carefully selected the right combinations of requirements which consist of importance variety to allow the requirements prioritization activity to happen.

TABLE I. Strength of the Agreement based on Cohen kappa

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

TABLE II. Symmetric Measures of GS and MosCoW Method

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.067	.055	2.033	.042
N of Valid Cases		31			

[a] Not assuming the null hypothesis

[b] Using the asymptotic standard error assuming the null hypothesis

TABLE III. SYMMETRIC MEASURES OF GS AND MCBRANK METHOD

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	**Kappa**	.600	.090	18.298	.000
N of Valid Cases		31			

[a] Not assuming the null hypothesis

[b] Using the asymptotic standard error assuming the null hypothesis

VI. CONCLUSION AND FUTURE WORK

An enhanced method is commonly understood to provide better results, but it is merely an assumption if scientific data does not prove the claim. Therefore, this paper presents an empirical investigation to show that an enhanced MoSCoW method named MCBRank improves software requirements prioritization. The empirical results showed that the requirements could be better prioritized closer to the Gold Standard and represent cumulative importance values from multiple stakeholders of a software system.

Further investigation can be done on various requirements elements and attributes that influence the prioritization for future work. Besides, techniques to be embedded for automation are worth exploring to expedite the prioritization process to improve performance. However, careful measures to include a variety of stakeholders' perspectives must be taken care of.

ACKNOWLEDGMENT

This paper is sponsored by Universiti Teknikal Malaysia Melaka through a research grant numbered PJP/2020/FTMK/PP/S01774.

REFERENCES

[1] Abd Elazim, K., Moawad, R. and Elfakharany, E., "A framework for requirements prioritization process in agile software development," In *Journal of Physics: Conference Series* vol. 1454, no. 1, pp. 1-11, 2020.

[2] Bukhsh, F.A., Bukhsh, Z.A. and Daneva, M., 2020. "A systematic literature review on requirement prioritization techniques and their empirical evaluation," *Computer Standards & Interfaces*, vol.69, pp.1-18, 2020.

[3] Asghar, A. R., Bhatti, S. N. Tabassum, A. Sultan, Z. and Abbas, R. "Role of requirements elicitation & prioritization to optimize quality in Scrum Agile Development," *International Journal of Advanced Computer Science and Applications (IJACSA)*. vol.7, no. 12, pp. 300-306, 2016.

[4] Ahmad, S. Asmai, S.A. "Measuring software requirements quality following negotiation through empirical study" *International Journal of Applied Engineering Research.* vol. 11 no. 6, pp. 4190-4196, 2016.

[5] Olaronke, I., Rhoda, I. and Ishaya, G. "An appraisal of software requirement prioritization techniques," *Asian Journal of Research in Computer Science*, vol. 1 no. 1, pp. 1–16, 2018.

[6] Gilb, T. and Maier, M. "Managing Priorities: A Key to Systematic Decision-Making." *In Proceedings of INCOSE International Symposium*. Wiley. 2005.

[7] Ma, Q. "The effectiveness of requirements prioritization techniques for a medium to a large number of requirements: A systematic literature review." *In Dissertation, Auckland University of Technology.* 2009.

[8] Berntsson Svensson, R. and Torkar, R., "Not All Requirements Prioritization Criteria Are Equal at All Times: A Quantitative Analysis," arXiv e-prints, pp.arXiv-2104, 2021.

[9] Jahan, M.S., Azam, F., Anwar, M.W, Amjad, A., and Ayub, K. "A novel approach for software requirement prioritization," *In Proceedings of 7th International Conference in Software Engineering Research and Innovation*. IEEE. pp. 1-7, 2019.

[10] Hujainah, F., Bakar, R.A., Alhroob, E., Al-haimi, B. and Nasser, A.B., "Interrelated Elements in Formulating an Efficient Requirements Prioritization Technique," *In 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* pp. 97-101, IEEE, 2020.

[11] Hujainah, F., Bakar, R.A., Nasser, A. B., Al-haimi, B. and Zamli, K. Z. "SRPTackle : A semi-automated requirements prioritisation technique for scalable requirements of software system projects," *Inf. Softw. Technol.*, vol. 131, 2020.

[12] Yaseen, M., Mustapha, A., Ibrahim, N., Rahman, A.U., Kamal, S.W. and Ijaz, A., "Importance of functional requirements prioritization: ODOO ERP as case study," *i-Manager's Journal on Software Engineering*, vol. 14, no. 1, 2020.

[13] Sasank, V.V.S., Prabha, B., Praveen, S. R. K., Hasane, S.K., Ahammad, G . Kumar, N.S. "Modern Driven Requirements Engineering with Requirements Prioritization in Software Testing" *Turkish Journal of Physiotherapy and Rehabilitation*. vol. 32, no. 2, pp. 3037–3043, 2021.

[14] Achimugu, P., Selamat, A., Ibrahim, R., Mahrin, M. N. "A systematic literature review of software requirements prioritization research," *Information and Software Technology*. vol. 56, no. 6, pp. 568–585, 2014.

[15] Miranda, E.,"Moscow Rules: A Quantitative Exposé," *International Conference on Agile Software Development*, pp. 19-34, 2022.

[16] Abdelazim, K., Moawad, R., and Elfakharany, E. "A Framework for Requirements Prioritization Process in Agile Software Development," *J. Phys. Conf. Ser.*, vol. 1454, no.1, 2020.

[17] Abdelazim, K., Moawad, R., and Elfakharany, E. "A Supporting Tool for Requirements Prioritization Process in Agile Software Development," *Futur. Comput. and Informatics J.*, vol. 5, no. 1, pp. 15–27, 2020.

[18] Tufail, H., Qasim, I., Masood, M. F., Tanvir, S. and Butt, W. H. "Towards the selection of Optimum Requirements Prioritization Technique: A Comparative Analysis," *5th Int. Conf. Inf. Manag. ICIM*, pp. 227–231, 2019.

[19] Saher, N., Baharom, F. and Romli, R., "A review of requirement prioritization techniques in Agile software development," *Knowl. Manag. Int. Conf.*, pp. 25–27, 2018.

[20] Jarzębowicz, A. and Sitko, N., "Agile requirements prioritization in practice: Results of an industrial survey," *Procedia Computer Science*, vol. 176, pp.3446-3455, 2020.

[21] Hudaib, A., Masadeh, R., Qasem, M.H. and Alzaqebah, A., "Requirements prioritization techniques comparison." *Modern Applied Science*, vol. 12, no. 2, p.62, 2018.

[22] Avesani, P., Ferrari, S. and Susi, A. "Case-based ranking for decision support systems," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2689, pp. 35–49, 2003.

[23] Yaseen, M., Mustapha, A., Rahman, A.U., Khan, S. and Kamal, W. "Importance of requirements prioritization in parallel developing software projects," *International Journal of Scientific Research in Computer Science and Engineering*, vol. 9, no. 2, pp.171-179, 2020.

[24] Gerogiannis, V.C., Fitsilis, P., Kakarontzas, G. and Born, C., "Handling vagueness and subjectivity in requirements prioritization" *In Proceedings of the 22nd Pan-Hellenic Conference on Informatics*, pp. 150-155, 2018.

[25] Ahmad, S. "Measuring the effectiveness of negotiation in software requirements engineering" Thesis of The University of Western Australia, 2012.

[26] Cohen, J. "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37-46, 1960.

[27] Landis, J. R. and Koch, G. G. Landis and Koch “Agreement of Categorical Data,” *Biometrics,* vol. 33, no. 1, pp. 159–174. 1977.

[28] Gerogiannis, V.C., Tsoni, E., Born, C. and Iatrellis, O. “Software Features prioritization based on Stakeholders’ Satisfaction/ Dissatisfaction and Hesitation” *In 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)* pp. 265-271, IEEE, 2020.

Breast Cancer Detection and Classification using Deep Learning Xception Algorithm

Basem S. Abunasser[1], Mohammed Rasheed J. AL-Hiealy[2], Ihab S. Zaqout[3], Samy S. Abu-Naser[4]
University Malaysia of Computer Science & Engineering (UNIMY), Cyberjaya, Malaysia[1, 2]
Faculty of Engineering and Information Technology, Al-Azhar University, Gaza, Palestine[3, 4]

***Abstract*—Breast Cancer (BC) is one of the leading cause of deaths worldwide. Approximately 10 million people pass away internationally from breast cancer in the year 2020. Breast Cancer is a fatal disease and very popular among women globally. It is ranked fourth among the fatal diseases of different cancers, for example colorectal cancer, cervical cancer, and brain tumors. Furthermore, the number of new cases of breast cancer is anticipated to upsurge by 70% in the next twenty years. Consequently, early detection and precise diagnosis of breast cancer plays an essential part in enhancing the diagnosis and improving the breast cancer survival rate of patients from 30 to 50%. Through the advances of technology in healthcare, deep learning takes a significant role in handling and inspecting a great number of X-ray, Magnetic Resonance Imaging (MRI), computed tomography (CT) images. The aim of this study is to propose a deep learning model to detect and classify breast cancers. Breast cancers has eight classes of cancers: benign adenosis, benign fibroadenoma, benign phyllodes tumor, benign tubular adenoma, malignant ductal carcinoma, malignant lobular carcinoma, malignant mucinous carcinoma, and malignant papillary carcinoma. The dataset was collected from Kaggle depository for breast cancer detection and classification. The measurement that was used in the evaluation of the proposed model includes: F1-score, recall, precision, accuracy. The proposed model was trained, validated and tested using the preprocessed dataset. The results showed that Precision was (97.60%), Recall (97.60%) and F1-Score (97.58%). This indicates that deep learning models are suitable for detecting and classifying breast cancers precisely.**

Keywords—Breast cancer; deep learning; xception

I. Introduction

Breast Cancer is considered one of the most common type of cancers and is taken as the second, to cause death in women worldwide. In 2002, it was the second most common cancer globally, by means of exceeding of one million new cases. Despite the enhancements in early detection and knowing of the molecular foundations of the biology of the breast cancer, nearly 30% of the patients with “early-stage” breast cancer have disease recurred.

Determining the most real and least toxic therapy, molecular and clinical features of the tumor in certain need of inspection. General treatment of breast cancer comprises hormonal agents, immunotherapy and cytotoxic. These medications are used in adjuvant, transitional, and neoadjuvant modes. Overall, systemic agents are vigorous at the start of management in 90% of main breast cancers and 50% of metastases. Though, after a flexible period of time, progress does follow. At the present, opposition to treatment is not only popular but anticipated [7, 9].

Diagnosing BC early is the most important features of its treatment. Amongst the diverse types of diagnosing technique of BC, imaging is the main diagnostic method that can make offer important data on the patients of BC. It was revealed that a few techniques of imaging e.g.: “Magnetic Resonance Imaging (MRI), Single-Photon Emission Computed Tomography (SPECT), mammography, Computerized Tomography (CT), and Positron Emission Tomography (PET)” can be put into use for the identifying and observing of the patients of BC in diverse stages. Moreover, techniques of imaging, and the use of biomarkers of biochemical e.g.: “proteins, mRNAs, DNA, and microRNAs” can be used as an innovative analytic and therapy gears for BC patients [12, 13 and 14].

Deep Learning (DL) as a type of Machine Learning method has its working mechanism inspiration from the way human brain neurons process information. The most basic element of the DL networks are small nodes known as artificial neurons , which are usually arranged in layers wherein each neuron has connections to every neuron in the subsequent layer via weighted connections. Recently, the rise of DL technique has encouraged different areas of study to solve complex problem or enhance performances of existing study using the new technology. Example of application of DL includes machine translation, speech recognition, sentiment analysis, image recognition, face recognition, signal processing, etc. [17].

The current trend of applying DL technique in medical applications has reaped great success with the potential of DL technology to perform faster analysis with higher accuracy when compared with human practitioners. To give an example, a notable study by Google on the diagnostic classification of diabetic retinopathy has shown remarkable performance that exceeds the capabilities of domain experts [22]. In addition, the application of transfer learning techniques can be seen in several studies. As opposed to training a model from scratch, transfer learning method allowed the use of weight s trained previously on a specific task to be reused as the starting point for a model on another task.

It comes with the benefits of shorter training time and is possible to deliver better results. CNN models such as Alex Net, VGG, ResNet, and Xception are the most dominant models applied in transfer learning approach. Generally, there are two ways of applying transfer learning technique; First, pre-trained model with weights trained on ImageNet dataset can be used as feature extractor; Second, fine-tuning of pre-trained model on a new problem. Study utilized transfer learning technology with Xception neural network to speed up the training of model for distinguishing subjects [6].

II. Literature Review

Many researchers have employed artificial intelligence, expert systems, and neural networks in the diagnosis of BC to increase the screening accuracy. Usually hospitals use x-rays for diagnosis of BC, but lately, hospitals have been using mammography images as a substitute of x-rays, due to their easiness in analyzing and studying through intelligent models, which have increased the efficiency and accuracy of BC diagnosis.

A number of models and methods were proposed for increasing the efficacy of the diagnosis of BC. These methods include: “Linear Regression (LR), Artificial Neural Network (ANN), K-Nearest Neighbors (KNN) search, Softmax Regression, and Support Vector Machine (SVM), and Convolutional Neural Network (CNN)”.

The authors in [3] and [18] collected their datasets from Kaggle depository. They proposed a prediction model to predict whether or not a person has breast cancer and to provide awareness or diagnosis about it. The authors of these studies made a comparisons using accuracy of each results of the SVM, random forest (RF), Naive Bayes classifier, and logistic regression on the dataset to deliver a precise model for breast cancer prediction. The outcome of their experiments indicated that the techniques of machine learning models that were applied in their studies predicted breast cancer disease with an accuracy between 52.63% and 98.24%.

In these studies, the authors propose new KNN models for the need for early diagnosis and precise diagnostic procedures that clinicians can use to classify whether cancer is benign or malignant. The main objective of their study was to compare the results of supervised learning classification algorithms and to combine these algorithms using a classification technique called voting. Voting was a grouping method because they could combine multiple models to achieve higher classification accuracy. The datasets were collected from the University of Wisconsin. [8] achieved 98.90%, [16] achieved an accuracy of 97.60%, [17] achieved 97.13%, [10] achieved 99.9%, [27] achieved 98.10% [28] achieved 98.23% and [6] achieved 83.45%.

In the following studies like [15, 21, 29, 30, 31] the authors argued that early detection and prevention can significantly reduce the chances of passing away. An important fact about breast cancer prognosis was to improve the likelihood of cancer recurrence. Their studies aimed to find the probability of breast cancer recurrence using different machine learning techniques like SVM. The authors presented new approaches in order to improve the accuracy of these models. Cancer patient data were collected from the Wisconsin Dataset of the UCI Machine Learning Repository. The dataset contains a total of 35 attributes in which they applied the Naive Bayes, C4.5 Decision Tree and SVM algorithms and measured their prediction accuracy. The efficient feature selection algorithm helped them improve the accuracy of each model by reducing some lower-order features. Not only are the contributions of these traits much lower, but their addition also misleads the classification algorithms. After careful selection of higher-order attributes, they significantly improved accuracy rate for all algorithms.

The study of [4, 10, 11, 24, 25, 26, 32] proposed method that uses CNNs which is a particular type of deep learning, feedforward network, and some preliminary experiments used the deep learning approach to classify breast cancer mammography images from BreaKHis, which is a public dataset. Method based on the extraction of image patches for training the CNN and the combination of these patches for final classification. All their convolutional network technique for categorizing screening mammograms reached good accuracies. The finest performance on an independent test set of digitized film mammograms from the Digital Database for Screening Mammography was 0.88%, (the sensitivity: 86.2%, the specificity: 80.2%).

The aim of the studies [1, 2, 5] was to develop methods for classifying cancers into specific prognostic categories based on gene expression signatures using artificial neural networks. ANN were trained to use small round blue cell tumors (SRBCTs) as a model. These cancers belong to four distinct diagnostic classes and often present diagnostic dilemmas in clinical practice. ANN properly classified all samples and recognized the genes most related to the classification. The experimental results suggested that the new strategies were able to improve the stability of the selection results as well as the sample classification accuracy. The new algorithms achieved accuracy of classification about 99%.

A. Previous Studies Summary

The previous studies aim was to detect if an image has breast cancer or not. The current study’s aim is to classify eight types of breast cancers. A summary was made of the studies discussed in the previous section in terms of the following criteria: Machine Learning methods used, programming language used, best result attained, best method, and source of the dataset used in Table I.

There are numerous studies and suggestions that put on to deal with breast cancer. Each suggestion or study has a different way of detecting or dealing with it. Some of the studies use methods that rely on image processing and the use of systems for this purpose and other studies that address the form of the breast and observe any changes in it. Some studies have enhanced the quality of algorithms that exist for detecting the disease. All these studies concentrate on the detection of whether breast cancer exists or not. None of these studies classified the eight different classes of the breast cancer. In the current study, the concentrate will be on the classification of the eight classes of breast cancers: “benign adenosis, benign fibroadenoma, benign phyllodes tumor, benign tubular adenoma, malignant ductal carcinoma, malignant papillary carcinoma, malignant mucinous carcinoma, and malignant lobular carcinoma”.

TABLE I. A SUMMARY OF THE PREVIOUS STUDIES BY METHODS, BEST METHOD, LANGUAGE, RESULT, DATASET USED

Reference	Methods Used	Best Methods	Language	Best Result	Data Provider
[1]	ANN	ANN	-	96.00	Private
[2]	ANN	ANN	Python	90.30	Kaggle
[3]	LR	LR	-	97.00	Wisconsin
[4]	CNN	CNN	Python	89.50	BreaKHis
[5]	ANN	ANN	-	99.00	Wisconsin
[6]	KNN	KNN	-	83.45	Wisconsin
[8]	KNN	KNN	-	98.90	FMC
[10]	CNN	CNN	Python	90.00	BreaKHis
[11]	CNN	CNN	Python	88.00	DDSM
[15]	SVM,KNN	SVM	-	99.00	Wisconsin
[16]	SVM,KNN,ANN	SVM	-	97.60	Wisconsin
[17]	SVM,KNN	KNN	-	97.13	Wisconsin
[18]	SVM,LR	LR	Python	98.24	Kaggle
[21]	SVM	SVM	-	92.30	Wisconsin
[22]	SVM,KNN	KNN	-	99.90	Wisconsin
[23]	SVM,ANN	SVM	-	98.82	Wisconsin
[24]	SVM,KNN,CNN	CNN	-	97.00	Wisconsin
[25]	SVM,KNN,CNN	CNN	Python	99.30	Wisconsin
[26]	SVM,ANN	ANN	Python	94.00	DDSM
[27]	SVM,KNN	SVM	Minitab	98.10	Wisconsin
[28]	SVM,KNN	SVM	WEKA	97.13	Wisconsin
[29]	SVM,KNN,LR,ANN	ANN	-	99.30	Wisconsin
[30]	SVM,KNN	SVM	SAE	98.90	Wisconsin
[31]	SVM,ANN	ANN	-	95.70	Private
[32]	KNN,LR,SVM,CNN	CNN	Python	97.20	Wisconsin

III. METHODOLOGY

The methodology that was used in the study consists of dataset collection, dataset preparation, dataset splitting, create the proposed model, model training, validating and testing as can be seen in Fig. 1.

A. Dataset

The dataset has been collected from Kaggle depository. The Breast "Cancer Histopathological Image Classification" (BreakHis) consists of 7909 microscopic images of breast cancer tissue gathered from 82 patients using various magnifying factors (40X, 100X, 200X, and 400X). It contains 2,480 benign and 5,429 malignant samples (700X460 pixels, three-channel RGB, eight-bit depth in each channel, PNG image format). This dataset was built in collaboration with the P&D Laboratory – Pathological Anatomy and Cytopathology, Parana, Brazil"

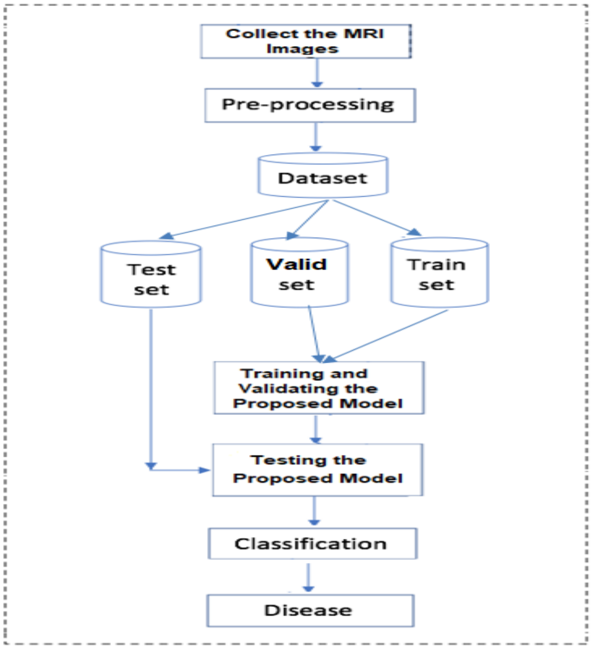


Fig. 1. Methodology Flowchart.

B. Data Perpetration

The number of images in the original BreakHis dataset has 7,909 images. This number of images is considered low and thus one can use new methods to generate more images to boost the dataset called Generative Adversarial Networks" (GANs). GAN is an algorithmic constructions that utilizes two NNs, fighting one in contradiction of the other (consequently the 'adversarial') to be able to produce new, fake instances of images that can pass as real images. They are very popular in image generation, voice generation, and video generation [19, 20, 23]. By using GAN the dataset was increased to 10,000 images. Each class of the eight classes has 1250 images. Sample of the eight classes are shown in Fig. 2.

C. Data Splitting

The dataset has been split into three datasets: training, validating, and testing datasets. The ratio of splitting is 60%, 20%, and 20% respectively.

D. Performance Measures

The most popular measurements were used in the performance of the proposed model: Accuracy is represented in (1), Precision is represented in (2), Recall is represented in (3) and F1-score is represented in (4).

$$\text{Accuracy} = (TP + TN) / (TP+TN+FP+FN) \quad (1)$$

$$\text{Precision} = (TP / (TP+FP) \quad (2)$$

$$\text{Recall} = (TP / (TP+FN) \quad (3)$$

$$\text{F1-score}= 2x(\text{Precision x Recall})/(\text{Precision}+\text{Recall}) \quad (4)$$

Where TP = True Positive, TN = True Negative

FP = False Positive, FN = False Negative

E. Proposed Model

A pre-trained Deep Learning model called Xception was proposed to be fine-tuned for the detection and classification of breast cancer. Xception model is considered to be among the beast pre-trained model of all classical Deep learning models due to its high accuracy in classifying the1000 natural images of ImageNet [17].

To be able to use a pre-trained model in the current dataset, it must be fine-tuned. In the current dataset, there are eight classes only. That means the top layer has to be removed from the Xception model which is called the classifier and to be replaced with the current classifier of the eight classes as illustrated in Fig. 3.

F. Model Training and Validating and Testing

The newly customized Xception model was trained with the prepared training dataset and cross-validating it with the validation dataset for 120 epochs. During the training, Learning Rate (0.0001), Batch Size (128), and the Optimizer is Adam were used. Furthermore, data augmentation was used during training to overcome the problem of overfitting. Fig. 4 and Fig. 5 shows the loss and accuracy of the training and validation of the xception model.

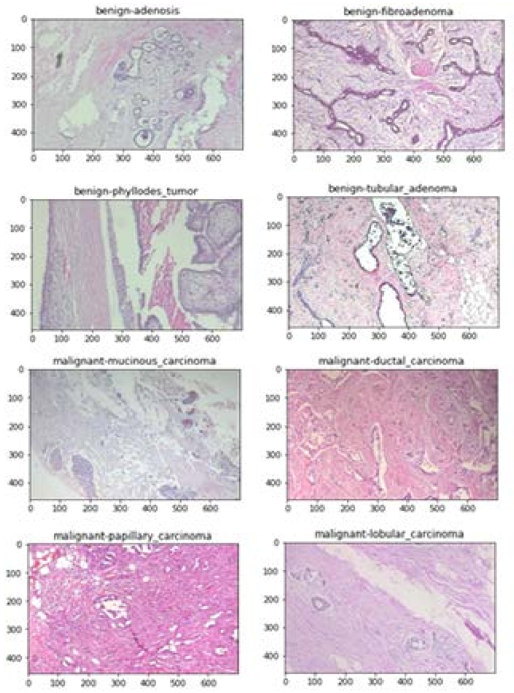

benign-adenosis
benign-fibroadenoma
benign-phyllodes_tumor
benign-tubular_adenoma
malignant-mucinous_carcinoma
malignant-ductal_carcinoma
malignant-papillary_carcinoma
malignant-lobular_carcinoma
0
100
200
300
400
500
600

Fig. 2. Sample of the Dataset with its 8 Different Classes.

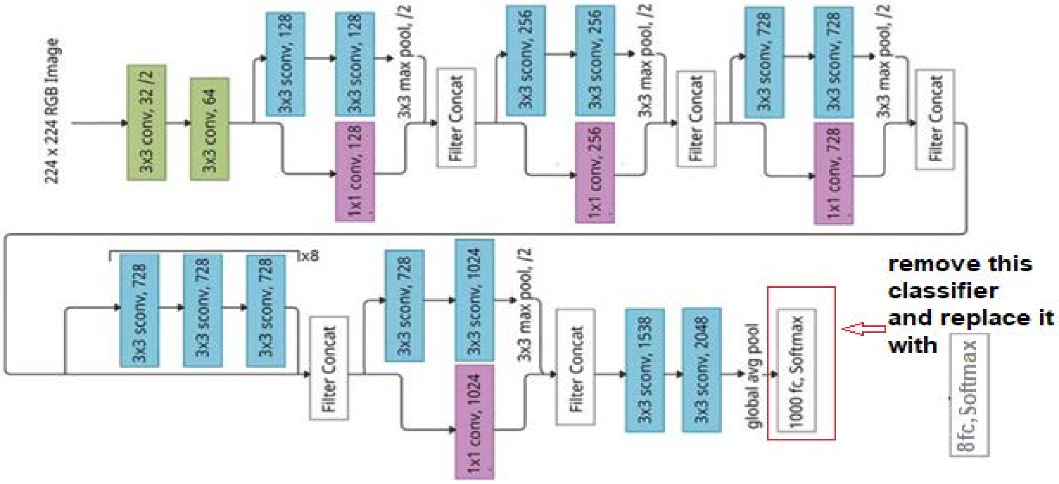


Fig. 3. Architecture of the Customized Xception Model.

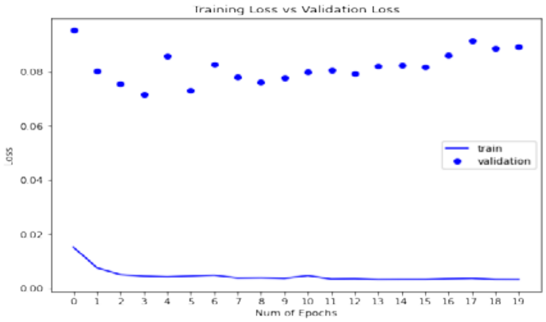


Fig. 4. Training and Validation Loss.

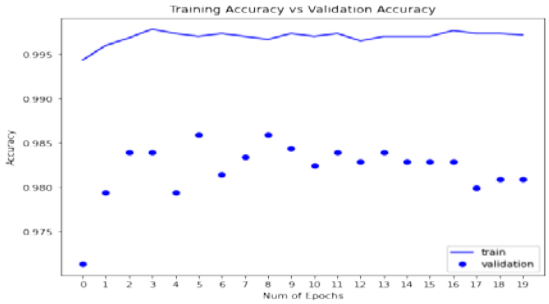


Fig. 5. Training and Validation Accuracy.

After the training was finished and validating the customized Xception model, it was tested with the testing dataset and the different performance measures were recorded.

IV. Result and Discussion

The model achieved Training Accuracy (99.78%), Validating Accuracy (98.59%) and Testing Accuracy (97.60%). In the customized model Training Loss was (0.00315), Validating Loss (0.07326), Testing Loss (0.09518). In terms of the time required for training and testing, the Xception model needed 2944 seconds for training and 5.32 seconds for testing.

Table II shows the precision, Recall, F1-Score of each class in the dataset in terms of the eight classes that the models are used for classifying: benign adenosis (BA), benign fibroadenoma (BF), benign phyllodes tumor (BPT), and benign tubular adenoma (BTA), malignant lobular carcinoma (MLC), malignant mucinous carcinoma (MMC), and malignant papillary carcinoma (MPC), malignant ductal carcinoma (MDC). The customized model achieved average Precision (97.60%), Recall (97.60%) and F1-Score (97.58%). Furthermore, the ROC Curve measure for each class in the dataset reached 100% as shown in Fig. 6.

TABLE II. Xception Precision, Recall, F1-Score of each Class in the Dataset

Measure	**BA**	**BF**	**BPT**	**BTA**	**MLC**	**MMC**	**MPC**	**MDC**	**Model AVG**
Precision	99.20%	97.63%	100.00%	98.01%	95.34%	94.16%	97.62%	98.81%	97.60%
Recall	99.60%	98.80%	98.80%	98.40%	90.00%	96.80%	98.40%	100.00%	97.60%
F1-Score	99.40%	98.21%	99.40%	98.20%	92.59%	95.46%	98.01%	99.40%	97.58%

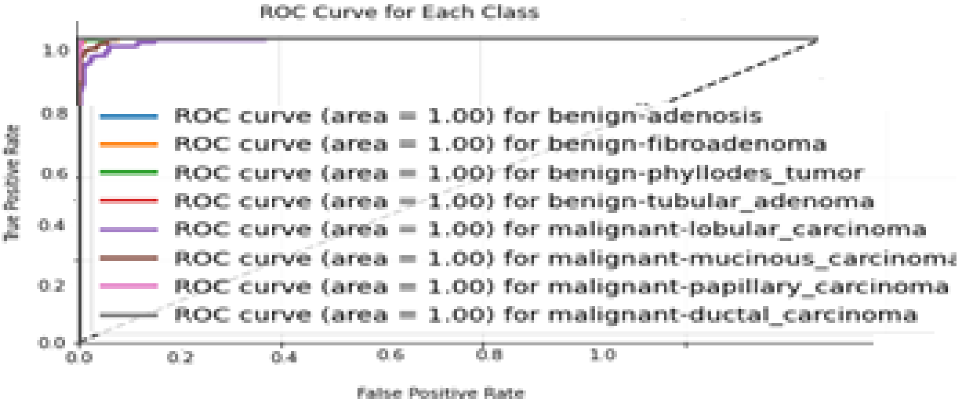


Fig. 6. ROC Curve for each Class in the Dataset.

V. Conclusion and Future Work

Breast Cancer is the top cause of deaths worldwide. About ten million people died globally from cancer in the year 2020. Breast Cancer is a fatal disease among women worldwide. With technical advances in healthcare, machine learning and deep learning play an important role in processing and analyzing a large number of medical images. The objective of this study is to propose a deep learning model for detecting classifying Breast Cancer. Xception model was used and customized to fit our current breast cancer eight classes dataset. The Dataset was collected from Kaggle and boosted using GAN. The dataset was split into three datasets: training, validating and testing. The customized Xception model was trained, validated, and tested. The mode achieved Precision (97.60%), Recall (97.60%) and F1-Score (97.58%).

In future work, other techniques of generating images will be used and the Xception model will be tested with newly generated images. Furthermore, other pre-trained model will be tested and compared with the current results.

References

[1] Westermann et al., "Classification and diagnostic prediction of pediatric cancers using gene expression profiling and artificial neural networks", *GBM Annual Fall meeting Halle 2020*, vol. 2020.

[2] S. Joo, Y. Yang, W. Moon and H. Kim, "Computer-Aided Diagnosis of Solid Breast Nodules: Use of an Artificial Neural Network Based on Multiple Sonographic Features", IEEE Transactions on Medical Imaging, vol. 23, no. 10, pp. 1292-1300, 2021.

[3] T. Kiyan and t. Yildirim, "breast cancer diagnosis using statistical neural networks", Istanbul university journal of electrical & electronics engineering, vol. 4, no. 2, 2021.

[4] F. Spanhol, L. Oliveira, C. Petitjean and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks", 2017.

[5] H. Salem, G. Attiya, N. El-Fishawy. Early diagnosis of breast cancer by gene expression profiles. Pattern Analysis and Applications. Vol. 20, no. 2, pp. 567–578, 2017.

[6] U. K. Kumar, M. B. S. Nikhil and K. Sumangali, "Prediction of breast cancer using voting classifier technique," 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2017, pp. 108-114.

[7] Albatish, I.M., Abu-Naser, S.S. Modeling and controlling smart traffic light system using a rule based system. Proceedings - 2019 International Conference on Promising Electronic Technologies, ICPET 2019, 2019, pp. 55–60.

[8] N. Ponraj, E. Jenifer., P. Poongodi, S. Manoharan. "Morphological operations for the mammogram image to increase the contrast for the efficient detection of breast cancer", European Journal of Scientific Research, (ISSN) 1450-216X, vol. 68, no.4, pp. 494-505, 2021.

[9] Naser, S. S. A. JEE-Tutor: An intelligent tutoring system for java expressions evaluation. Information Technology Journal, 2008, Vol. 7, No. 3, 528-532.

[10] B. Gayathri, C. Sumathi, T. Santhanam, "Breast cancer diagnosis using machine learning algorithm a survey". International Journal of Distributed and Parallel Systems, vol. 4, no. 3, pp. 39-50, 2017.

[11] L. Shen, L. Margolies, J. Rothstein. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. Sci Rep vol. 9, no.1, pp. 24-35, 2019.

[12] World health organization https://www.who.int/.

[13] Abu-Naser, S.S., El-Hissi H., Abu-Rass, M., & El-khozondar, N. An expert system for endocrine diagnosis and treatments using JESS. Journal of Artificial Intelligence, 2010, Vol. 3, No. 4, pp. 239-251.

[14] Abu Naser, S.S. Evaluating the effectiveness of the CPP-Tutor, an intelligent tutoring system for students learning to program in C++. Journal of Applied Sciences Research, 2009, Vol. 5, No. 1, pp. 109-114.

[15] A. F. Agarap. On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. ArXiv, abs/1711.07831, 2018.

[16] D. R. Pavithra, S. Preethi, & A. K. SriRakshitha. Breast Cancer Classification using the Supervised Learning Algorithms. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1492-1498, 2021.

[17] H. Asri, H. Mousannif, H. Moatassime, & T. Noël. Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. ANT/SEIT, 2017.

[18] A. M. Gonzalez-Angulo, F. Morales-Vasquez, G. N. Hortobagyi. Overview of resistance to systemic therapy in patients with breast cancer. Adv Exp Med Biol. 2017; vol. 608, pp. 1-22.

[19] Saleh, A., Sukaik, R., Abu-Naser, S.S. Brain tumor classification using deep learning. Proceedings - 2020 International Conference on Assistive and Rehabilitation Technologies, iCareTech 2020, 2020, pp. 131–136, 9328072.

[20] Arqawi, S., Atieh, K.A.F.T., Shobaki, M.J.A.L., Abu-Naser, S.S., Abu Abdulla, A.A.M. Integration of the dimensions of computerized health information systems and their role in improving administrative performance in Al-Shifa medical complex, Journal of Theoretical and Applied Information Technologythis link is disabled, 2020, Vol. 98, No. 6, pp. 1087–1119.

[21] A. I. Pritom, M. A. R. Munshi, S. A. Sabab and S. Shihab, "Predicting breast cancer recurrence using effective classification and feature selection technique," 2016 19th International Conference on Computer and Information Technology (ICCIT), 2017, pp. 310-314.

[22] A. Joshi, D. A. Mehta. Comparative Analysis of Various Machine Learning Techniques for Diagnosis of Breast Cancer. International Journal on Emerging Technologies, vol. 8, (NCETST-2017), pp. 522–526, 2017.

[23] A. D. Omondiagbe, V. Shanmugam, S. S. Amandeep. Machine Learning Classification Techniques for Breast Cancer Diagnosis, vol. 5, no. 3, 2019.

[24] K.Wadkar, P. Pathak, N. Wagh. Breast cancer detection using Ann network and performance analysis with SVM. International journal of computer engineering and technology, vol. 10, no. 3, 2019, https://doi.org/10.34218/ijcet.10.3.2019.009.

[25] M. TIWARI, R. Bharuka, P. Shah, R. Lokare. Breast Cancer Prediction Using Deep Learning and Machine Learning Techniques. SSRN Electronic Journal. Published. https://doi.org/10.2139/ssrn.3558786, 2020.

[26] D. A. Ragab, M. Sharkas, S. Marshall, J. Ren. Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ, vol. 7, no. 6, pp. 20-31, 2019. https://doi.org/10.7717/peerj.6201.

[27] M. F. Ak. A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications. Healthcare, vol. 8, no. 2, pp.1-11, 2020.

[28] H. Mousannif, H. Asri, H. A. Moatassime, T. Noel. Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. Procedia Computer Science, vol. 83, pp. 1064–1069, 2017.

[29] R. MurtiRawat, S. Panchal, V. K. Singh and Y. Panchal. Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning. International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 534-540, doi: 10.1109/ICESC48915.2020.9155783.

[30] D. Selvathi and A. A. Poornila. Performance Analysis of Various Classifiers on Deep Learning Network for Breast Cancer Detection. International Conference on Signal Processing and Communication (ICSPC), pp. 359-363, IEEE 2017.

[31] A. LG, E. AT. Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence. Journal of Health & Medical Informatics, vol. 04, no. 02.2018.

[32] M. A. Naji, S. E. Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouhahid, O. Debauche. Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis. Procedia Computer Science, vol. 191, pp. 487–492, 2021.

Improved POS Tagging Model for Malay Twitter Data based on Machine Learning Algorithm

Siti Noor Allia Noor Ariffin[1], Sabrina Tiun[2]
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

***Abstract*—Twitter is a popular social media platform in Malaysia that allows for 280-character microblogging. Almost everything that happens in a single day is tweeted by users. Because of the popularity of Twitter, most Malaysians use it daily, providing researchers and developers with a wealth of data on Malaysian users. This paper explains why and how this study chose to create a new Malay Twitter corpus, Malay Part-of-Speech (POS) tags, and a Malay POS tagger model. The goal of this paper is to improve existing Malay POS tags so that they are more compatible with the newly created Malay Twitter corpus, as well as to build a POS tagging model specifically tailored for Malay Twitter data using various machine learning algorithms. For instance, Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), and K-Nearest Neighbor (KNN) classifiers. This study's data was gathered by using Twitter's Advanced Search function and relevant and related keywords associated with informal Malay. The data was fed into machine learning algorithms after several stages of processing to serve as the training and testing corpus. The evaluation and analysis of the developed Malay POS tagger model show that the SVM classifier, as well as the newly proposed Malay POS tags, is the best machine learning algorithm for Malay Twitter data. Furthermore, the prediction accuracy and POS tagging results show that this research outperformed a comparable previous study, indicating that the Malay POS tagger model and its POS were successfully improved.**

Keywords*—*Informal Malay; Malay Twitter corpus; Malay POS tagging; Malay POS tagger model, Malay social media texts; Malay POS machine learning

I. Introduction

In general, POS is a classification system for words that classifies them according to their usage and function in sentences [1]. Similarly, a POS tagger is a component of software that reads the text in multiple languages and assigns appropriate words to each word (or other tokens) in the text. Malay has four leading POS tags, namely nouns, verbs, adjectives, and word tasks [2]; however, these leading POS tags are more suitable for tagging standard Malay text than informal Malay text such as Malay Twitter data.

Twitter is a microblogging service that combines social networking websites and instant messaging technologies to create a network of users from all walks of life who can communicate throughout the day via 280-character [3][4] short messages called tweets. Additionally, users can use Twitter's Trending features to follow specific topics. Tweets can range from jokes to current events to dinner plans, yet they cannot exceed Twitter's characters limit [5].

Twitter's Application Programming Interface (API) simplifies the process of collecting Twitter data. However, Twitter imposed specific fees and requirements for data access. Clearly, according to [6], collecting tweets using a standard Twitter API account is limited to the most recent seven days of data, and collecting and accessing tweets older than those seven days requires a premium Twitter API account efficiently cost hundreds of dollars. On top of that, Twitter provides an Advanced Search feature that enables users to refine search results based on date ranges, people, and more. Hence, this study collects tweets written in informal Malay, incorporates various dialects, conversational slang languages, and mixed languages via Twitter's Advanced Search feature [1][6][7], rather than the costly and limited API [6]. However, this data collection limited to the words contained in tweets due to other tweet features, such as user information (full name & username), hashtags, URLs, and timestamps, are considered superfluous. Hence, this study purposefully ignored and omitted it.

Malay (ISO 639-3; MSA) is a formal language spoken by Malaysians, Indonesians, Singaporeans, and Bruneians of all races. On the contrary, informal Malay is a dialect of Malay that Malaysians use in everyday conversation. Informal Malay encompasses a diverse range of informal terms, such as accent (or dialect) words, slang, titles (e.g., hang, mek), sounds (such as words written to express sounds like laughter, cat sounds, and knocking sounds), and mixed languages. Firstly, the term 'regional dialect or language' refers to a group of people who speak a country state's language, resulting in word variation. In comparison, slang is understood by a minuscule percentage of the population. Following that, 'mixed language' refers to the simultaneous use of a foreign language and Malay. For instance, when users write on social media to share feedback, express an opinion or stories, they frequently use every day conversational language to convey a friendly, casual, and easy-going image to other users.

Moreover, numerous research papers have been published in recent years on the Malay Twitter corpus and the prediction and POS tagging of Malay Twitter data; however, there has been a dearth of information in various areas, which requires improvement. For instance, Malay Twitter data normalization techniques [5], Malay POS tagger explicitly tailored for such data [1], and supervised machine learning algorithms that best suit the mentioned data. According to [8], one way for improving the quality of language processing on social media data is to automate non-standard terms to their corresponding standard tokens using normalization methods. Hence, this study chose to utilize [5]'s improvised normalization techniques for Malay Twitter data, dubbed Malay Text Normalizer. Furthermore, their Malay Text Normalizer demonstrated acceptable performance with POS tagging on a normalized tweets test corpus.

Besides, Malay Twitter data can be tagged with POS in numerous ways. Using supervised machine learning algorithms to build a POS tagging model is one of the most widely used methods. According to [1], developing a Malay POS tagging model specifically for Malay Twitter data is difficult due to the presence of dialects, grammatical and typographical errors, and abbreviations. Nevertheless, they have successfully developed a Malay POS tagger that can tag Malay Twitter data using a supervised machine learning algorithm, QTAG, a language-independent probabilistic tagger by [9]. In addition, their Malay QTAG tagger produces exceptional results on both normalized and unnormalized test corpora. Therefore, in this study several different supervised machine learning algorithms such as SVM, NB, DT, and KNN classifiers will be employ. This technique has been applied to a wide variety of research fields. The primary reason for using these supervised machine learning algorithms in this study is that it has demonstrated satisfactory results in other languages and is underutilized in informal Malay Twitter data. Thus, it is critical to investigate how well these supervised machine learning algorithms perform on such data.

This study aims to enhance existing Malay POS tags by [1] to make them more compatible with the newly created Malay Twitter corpus and develop a POS tagging model specifically for the Malay Twitter corpus using the previously mentioned machine learning algorithms. Throughout this paper, this study has successfully made several significant contributions, such as collecting and extracting Malay Twitter data by employing informal Malay terms as keywords, tagging the Malay Twitter corpus with newly proposed improvised Malay POS tags, and analyzing the corpus data with the newly developed data Malay POS tagging model. Nonetheless, until the data's copyright is enforced, this Malay Twitter data collection will be inaccessible to the public or future research.

The following is how the rest of this paper is organized: Section II summarizes relevant works, Section III discusses the methodology of this study, Section IV discusses the discussion of this study, Section V presents the results and analysis, and Section VI summarizes the study as well as several suggested future works.

II. Related Work

Malay POS tags are a type of Malay POS used to tag words in this language. As stated previously, Malay has four primary POS tags; however, this POS tag is insufficient for tagging words in the Malay Twitter corpus. For this reason, this study decided to create new Malay POS tags by referencing Malay formulas and grammar by [2]. Furthermore, this study modified the newly created POS tags to meet the Malay Twitter data criteria by comparing the word classes discovered in [2] and [10], as well as some previous findings on social media texts [1].

Additionally, a study by [11] created several new POS tags to meet their research requirements. This study discovered that several of [11]'s newly developed POS tags are suitable for categorizing words in the Malay Twitter corpus. Consider the FOR and NEG POS tags, for example. The FOR POS tag is used to classify words in foreign languages found in the study corpus. Similarly, the NEG POS tag identifies words with negative connotations, such as those that refer to swearing. Therefore, this study chose to use [11]'s two newly developed POS tags, namely FOR and NEG POS tags, as the Malay Twitter data writing demonstrates that Malaysians enjoy combining words from multiple languages in tweets and using negative words to express emotions or disapproval of situations. In addition, removing a foreign language from the corpus alters the author's intended meaning and the structure of the tweets' writing. This change will result in an error when auto annotations and annotators attempt to tag the POS tags. In that case, this study will not exclude foreign language terms, slang terms, or informal Malay expressions that adhere to this principle—instead, a unique POS tag designed for this type of word tagging. Moreover, by referencing to [11], this study generated several new Malay POS tags, namely LD, SL, GL, and BY POS tags. Firstly, the LD POS tag is used to indicate words with an accent (or dialect). Secondly, the SL POS tag is used to indicate slang language. Following that, the GL POS tag is used to indicate words referred to nicknames, and lastly, the BY POS tag used to indicate words that express sounds. Besides, this study also combined two POS tags that contained the same word into a single POS tag, such as the GDT-KTY POS tag, which comprises the POS tag for self-query pronouns (kata ganti nama diri tanya) and the POS tag for query word (kata tanya). Finally, another newly developed Malay POS tag was created by combining several prominent POS tags with sub-POS tags for shortened words, accents (or dialects), and negative particles. For instance, the term 'iols' is an abbreviation for a foreign language term. Therefore, the appropriate POS tag for this term combines the foreign language primary POS tag (FOR) and the shortened word sub-POS tags (KEP).

Malaysians, particularly teenagers, are incredibly inventive in writing and developing new words [12]. As a result, Malay social media text, such as Malay tweets, is saturated with informal Malay and peppered with mixed-language phrases and derogatory terms. POS tagging is exceptionally complicated, time-consuming, and energy-intensive for this type of corpus. If a word is not suitable to any existing POS tags, researchers will need to find alternative initiatives to either remove the word or tag it with any existing POS tags that they deem appropriate. This technique, however, is impractical due to the researchers' manipulation of study data. Thus, this study took the initiative to create new Malay POS tags that are compatible with Malay Twitter data to avoid confusion and expedite the POS tagging process; this study not only added and used the FOR and NEG POS tags from [11] and created several new Malay POS tags, but also added six additional Malay POS tags from [1]. With these numerous newly crafted Malay POS tags, this study's total number of POS tags has increased to 45 tags. As stated earlier, Malay POS tags were explicitly created to tag words in Malay Twitter data, saving researchers and annotators time to tag words with the correct POS tags.

This study conducted prediction and Malay POS tagging using the Malay POS tagger model developed based on four different machine learning algorithms from similar past studies [1], namely SVM, NB, DT, and KNN classifiers. Firstly, SVM is a widely used classifier in this type of research using machine learning algorithms, as it can accurately predict and tag POS tags [13]. Additionally, SVM predicts the POS tags for unknown words [14][15] and is considered one of the most efficient and accurate classification techniques for POS tagging [16]. According to [14], this classifier is based on sub-words, contextual information, and environmental context tagger, and [16] noted that this classifier could be used to classify sentences, ambiguity classes, and word length. Secondly, the NB classifier generates documents using the Bayes rule theory. NB classifier is extensively used in research related to predicting and tagging POS tags [17], as it can estimate the probability of each word feature before building the classification model. Five classification techniques are included in the NB class: Gaussian, Multinomial, Complement, Bernoulli, and Categorical. Following that, DT is a highly simple-to-understand and interprets machine learning class. DT works by developing predictive models for target variables. This classifier requires only a tiny amount of configuration data to operate and supports two distinct data types: numeric and categorical. DT can be validated using statistical tests and perform well, even if the data model occasionally rejects the resulting prediction. This classifier can overcome the information breakdown problem by obtaining accurate estimates of the probability of change [18] and can be used to tag unknown words with POS tags based on their word endings, word forms, and context information. Lastly, because the KNN classifier relies on POS tags associated with the test corpus, it does not generate a clear declaration representation; however, it estimated the new word POS tag by comparing it to words in the training set.

Generally, this study evaluated and quantified the prediction accuracy and POS tagging using these four-evaluation metrics: precision, recall, F1-score, and accuracy. Firstly, the precision evaluation metrics quantify machine learning algorithms' ability to avoid predicting and tagging negative samples as positive. Secondly, the recall evaluation metric assesses a machine learning algorithm's ability to re-identify positive samples. These terms refer to whether the predictions made by such machine learning algorithms are appropriate for external assessment or not. Following that the F1-score evaluation metrics can be interpreted as a weighted harmonic mean for precision and recall evaluation metrics. Furthermore, this study discovered that Malay social media text's prediction accuracy and POS tagging could be evaluated and quantified using all four-evaluation metrics. In addition, this study also discovered that the computation value for each evaluation metric could be easily generated via the classification report's evaluation metrics [19]. Classification reports are evaluation metrics that present computation values for each of the four-evaluation metrics in a classification report table. Thus, this study chose to evaluate and quantify the prediction accuracy and POS tagging using the sklearn library's evaluation metrics for classification reports [20].

III. Methodology

This study aims to improve existing Malay POS tags to fit the new Malay Twitter corpus better and develop a tagging model tailored to the new Malay Twitter corpus using the previously mentioned machine learning algorithms. The proposed methodology entails data collection and pre-processing as shown in Fig. 1. As stated previously, this study's data was compiled using Twitter's Advanced Search feature. It conducts the search using the provided keywords. Pre-processing of data includes data normalization and annotation. After that, vectorization is then applied to the data, preparing it for use by the machine learning algorithms: SVM, NB, DT, and KNN classifiers. This study vectorizes the data using the sklearn library's vectorization techniques [20], converting them to numbers, as machine learning algorithms operate exclusively on numbers. Then, to effectively use these algorithms, they must be trained to extract both word characteristics and their POS from the data. This procedure generates a model upon which Malay tweets can be automatically tagged with the Malay POS tags based solely on their context. The following sections go into details about each process.

A. Data Collection

This study gathered data manually by focusing on keywords associated with informal Malay and restricting the date ranges to February 2019. The keywords were selected following a review of the literature on informal Malay and structure. Table I contains a sample of keywords derived from previous studies and used to collect data. As stated earlier, the standard method of collecting Twitter data is through their API, enabling developers and researchers to collect data quickly; however, the API comes with a slew of restrictions, including a seven-day limit on tweets and a limit on the number of requests made to the Twitter server [6]. As a result, this study chose to manually collect data using Twitter's Advanced Search feature, rendering the limitations mentioned previously obsolete.

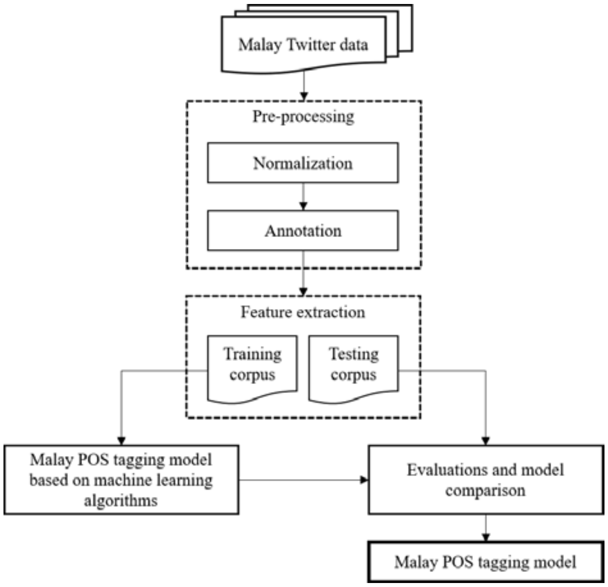


Fig. 1. Architecture of the Malay POS Tagging Model.

TABLE I. THE FOLLOWING ARE SOME OF THE KEYWORDS ASSOCIATED WITH SLANG TERMS [17] USED TO COLLECT DATA

Slang Word	Meaning in Malay	Meaning in English	Slang Form
bro	*awak, kamu, abang*	you, brother	Pronouns
gua	*saya, aku*	me	Pronouns
korang	*awak semua, kamu semua*	all of you, you all	Pronouns
pastu	*selepas itu*	after that	Expression
kekadang	*kadang-kadang*	sometimes	Expression
tetibe	*tiba-tiba*	out of the blue, suddenly	Expression
heran	*hairan*	astonished	Expression
takpe	*tidak mengapa*	it's okay	Expression
takleh	*tidak boleh*	can't	Expression
citer	*cerita*	story, tale	Expression
sesama	*bersama-sama*	together	Expression
jeles	*cemburu*	jealous	Foreign language
terer	*pandai*	clever	Foreign language
kaler	*warna*	colour, tone	Foreign language
hensem	*kacak*	handsome	Foreign language

This study focuses entirely on a single criterion for tweet inclusion: tweets must be written in informal Malay. As noted previously, informal Malay is replete with colloquial terms such as dialect, slang, titles, sounds, and mixed languages. Eventually, this study used only keywords derived from previous research findings to ensure that the chosen tweets were appropriate and accurately reflected this study's objective. On the contrary, there are currently no exclusion criteria for tweets as this study compiled a list of all tweets returned in response to the keywords entered. However, as explained earlier, several additional tweet features were purposefully overlooked and omitted, as the study's goal is to collect only informal Malay text. Therefore, this study ignores all other characteristics to focus exclusively on textual characteristics and the frequency with which informal Malay terms were used in social media texts.

B. Data Pre-processing

As previously stated, the data for this study will be pre-processed using two well-known pre-processing steps: data normalization and annotation. The pre-processing of the data begins with data normalization. The data normalization process strips Malay Twitter data of all ambiguous signs, symbols, and spellings. This normalization process is required for accurate POS tagging; however, accurate POS tagging is only possible after informal terms in the data are converted to their standard form (spellings). Therefore, this study's data were normalized using Malay Text Normalizer, a rule-based normalizer designed to normalize only Malay, Romanized Arabic, and English words [5] in the corpus. Subsequently, in the next pre-processing steps: data annotation, the normalized data will be annotated with POS. The data annotation process needs to ensure that each word in the data is appropriately tagged with the correct POS tags. For this reason, this study's data were manually annotated [1] using the newly proposed Malay POS tags. The final dataset contains 1,791 tweets in various languages relevant to informal Malay and related to the previously mentioned keywords.

C. Feature Extraction and Model Training

Following data collection and pre-processing, the final dataset is divided into two corpora: training and testing. The training corpus contains 70% of the data from the final dataset and is used to train these machine learning algorithms on extracting valuable features from words and their contexts using [21] features extraction method. This study extracted ten significant features, including the preceding and following words, the prefix of each word (limited to the first three alphabets of the word), the suffix of each word (limited to the last three alphabets of the word), the word's length, and the presence of a digit in the word. These features were extracted to aid the POS tagging model in automatically assign the correct POS to each word in the corpus, based on its context alone.

Additionally, as previously explained, the training corpus is vectorized before being fed into machine learning algorithms. Vectorization converts data from a textual to a numerical format, as these algorithms only work with numbers. Hence, this study leverages [20]'s sklearn library by vectorizing the training corpus using its vectorization technique. After that, all machine learning classifiers are trained using this vectorized training corpus by incorporating it into the classifier's code in the sklearn library. The sklearn library by [20] used in this study because it includes all necessary algorithms [22], such as machine learning classifiers and vectorization techniques, and evaluation methods. Besides, the library is simple to use, as the algorithms are invoked via the provided code. Finally, the accuracy of these machine learning classifiers will be evaluated using the sklearn library's evaluation metrics for classification reports and tested using the testing corpus, which contains the remaining 30% of the final dataset's data. Table II contains a summary of the features used in this phase.

TABLE II. LIST OF FEATURES USED IN THIS STUDY

Categories	Feature Description
Word Prefixes & Suffixes	Prefix 1 (first letter) Prefix 2 (first two letters) Prefix 3 (first three letters) Suffix 1 (last letter) Suffix 2 (last two letters) Suffix 3 (last three letters)
Token (Word) Context	The previous word Next word Word length Does the word contain digits

The features listed in Table II were chosen because:

1) According to [21], the word before it and its POS tag serves as a guideline for identifying the appropriate POS tag for the word after it.

2) Author in [21] also stated that the presented algorithm tags the current word with the appropriate POS tags based on the word information and surrounding POS tags.

3) The letter at the beginning and end of each word refers to the letter's size and position in the word. This feature, according to [23], has the potential to influence the effectiveness of POS tagging.

4) The length of each word is a binary feature that determines its size [23], [24].

5) The author in [23] defines digit features as features based on the presence of digits (numbers) and symbols in words.

6) According to [23] and [24], the context of words in a sentence influences the tagging value of unigrams in the corpus.

IV. Discussion

This study gathered tweets written in informal Malay from numerous Malaysian users through relevant and related keywords associated with informal Malay language. The keywords searched was done by using the same method used by [1][6][7] which is through Advanced Search Twitter function located at the Twitter main homepage. The collected data then was normalized using a Malay normalizer by [5] and annotated with the newly created Malay POS tags in the pre-processing process. Additionally, this study extensively used the sklearn library code by [20] to implement the classifiers algorithm, vectorization techniques, and evaluation method. Following that, this study vectorized the data to prepare it for the machine learning algorithms as the classifiers can only read data in numerical format. Finally, this study used the sklearn library [20] to develop a Malay POS tagger model based on these machine learning algorithms, namely SVM, NB, DT, and KNN, and evaluate each classifier's results.

V. Results and Analysis

This study developed a Malay POS tagger models using a training corpus and machine learning classifiers code from the sklearn library in previous sections. Therefore, this section will evaluate the developed Malay POS tagger models using the test corpus to determine its functionality. The evaluation of POS tagging by Malay POS tagger models employs metrics for classification reports from the sklearn library to determine the accuracy of POS tagging across all four machine learning algorithms. Table III summarizes the algorithms' evaluation results using the prepared testing corpus and the details mentioned previously.

Based on the table above, this study discovered that the SVM classifier achieved a relatively high predictive accuracy and Malay POS tagging at 94%, and the DT classifier had the second-highest score at a rate of 93%. Besides, the SVM and DT classifiers also rated the highest POS tagging evaluation results on the same scoring metric, i.e., F1-score (0.92 & 0.89). In other words, this evaluation demonstrates that both classifiers generate nearly identical predictive outcomes and POS tagging. The reason for this is that while the SVM classifier is computationally simple to implement, it has a high computational cost [14][15], whereas the DT classifier is capable of handling sparse data problems and still produces the best results when tested on small size data [25][26].

TABLE III. The Following Summarized the Classification Reports for POS Tagging Generated by All Four Machine Learning Classifiers

Machine Learning Classifier	**Classification Reports**			
	Precision	***Recall***	***F1-score***	***Accuracy***
Support Vector Machine (SVM)	0.93	0.91	0.92	0.95
Naïve Bayes (NB)	0.74	0.57	0.60	0.85
Decision Tree (DT)	0.89	0.88	0.89	0.93
K-Nearest Neighbor (KNN)	0.85	0.87	0.85	0.90

This study examined the algorithms' functionality using extensive data collection, resulting in up to 95% accuracy. Based on this result, the developed Malay POS tagging model is ready to predict and tag Malay Twitter data using the newly created Malay POS tags. The evaluation and analysis results indicate that the SVM classifier is the optimal machine learning algorithm for Malay Twitter data and that the Malay POS tags used are also optimal for such data. Furthermore, this study's prediction accuracy and POS tagging score are higher than those of a similar previous study, indicating that this study successfully improved the Malay POS tagger model and its POS.

A. Model Comparisons

As stated earlier, this study was based on an earlier study by [1]. The author in [1] investigated the prediction and tagging of Malay POS by developing Malay POS tagger models based on the Hidden Markov Model (HMM) trigram machine learning algorithm, the QTAG model. As a result, this study took the initiative to conduct comparable studies using various machine learning algorithms, including SVM, NB, DT, and KNN classifiers. The purpose of this study was to determine which machine learning algorithms are best suited for processing Malay POS predictions and tagging, particularly Malay Twitter data.

While both studies used the same data type, the pre-processing of the data was significantly different. For instance, this study normalized the data using the Malay Text Normalizer by [5], whereas [1] used only a few simple data pre-processing steps such as the removal of punctuation marks, symbols, and numbers. Following that, the Malay POS tags used in the annotation process of the two studies are distinct, as each study developed its own Malay POS tags that correspond to the content of its study data. Finally, by comparing the prediction accuracy and POS tagging results from these two studies, the results obtained by this study were significantly higher, with a difference of 0.29% from [1]. Eventually, the results and analysis demonstrate that the SVM classifier is the optimal machine learning algorithm for predicting and tagging Malay POS in Malay Twitter data and that the newly proposed Malay POS tags are appropriate for use as POS in such data. The distinction between this study and [1] is summarized in Table IV.

TABLE IV. THE FOLLOWING TABLE COMPARES [1] TO THIS STUDY

Studies	[1]'s study	This study
Objectives	Malay POS tagger	
Corpus	Informal Malay tweets	
Technique	Machine learning	
Algorithms	HMM (QTAG)	SVM, NB, DT, & KNN
POS Size	38	45
No. of Tweets	300	1,791
No. of Words	5,513	38,714
Accuracy	94.60%	95.00%

VI. CONCLUSION AND FUTURE WORK

In general, this study aims to improve on [1], work by developing a POS tagging model tailored to Malay Twitter data using a variety of machine learning algorithms, including SVM, NB, DT, and KNN classifiers. The collected data, as well as the newly developed Malay POS tags and Malay POS tagger model, are expected to be of assistance to researchers and developers, particularly those with expertise in informal Malay and related Natural Language Processing fields. This study can be improved further by collecting more non-standard Malay words in the training corpus, preferably more than 10,000 tweets specific to a specific domain such as food or health, rather than general domain. This enhancement is proposed to ensure that the training corpus contains enough data for the subsequent processing process.

ACKNOWLEDGMENT

This research work is funded by The Ministry of Higher Education Malaysia under research grant code: FRGS/1/2020/ICT02/UKM/02/1.

REFERENCES

[1] Ariffin, S. N. A. N., & Tiun, S. "Part-of-Speech Tagger for Malay Social Media Texts". GEMA Online® Journal of Language Studies, 18(4), 2018.

[2] Safiah, K. N., Onn, F. M., Musa, H. H., & Mahmood, A. H. "Tatabahasa Dewan Edisi Ketiga". Kuala Lumpur: Dewan Bahasa dan Pustaka, 2010.

[3] Meftah, S., & Semmar, N. "A neural network model for part-of-speech tagging of social media texts". In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), May 2018.

[4] Kumar, P., & Gruzd, A. "Social Media for Informal Learning: a Case of# Twitterstorians". In Proceedings of the 52nd Hawaii International Conference on System Sciences, January 2019.

[5] Ariffin, S. N. A. N., & Tiun, S. "Rule-based text normalization for Malay social media texts". International Journal of Advanced Computer Science and Applications, 11(10), 2020.

[6] Feizollah, A., Ainin, S., Anuar, N. B., Abdullah, N. A. B., & Hazim, M. "Halal products on Twitter: Data extraction and sentiment analysis using a stack of deep learning algorithms". IEEE Access, 7, 83354-83362, 2019.

[7] Izazi, Z. Z., & Tengku-Sepora, T. M. "Slangs on Social Media: Variations among Malay Language Users on Twitter". Pertanika Journal of Social Sciences & Humanities, 28(1), 2020.

[8] Li, C., & Liu, Y. "Joint POS tagging and text normalization for informal text". In Twenty-Fourth International Joint Conference on Artificial Intelligence, June 2015.

[9] Tufis, D., & Mason, O. "Tagging Romanian texts: a case study for qtag, a language-independent probabilistic tagger". In Proceedings of the First International Conference on Language Resources and Evaluation (LREC) (Vol. 1, No. 589-596, p. 143), May 1998.

[10] Othman, A., & Karim, N. S. "Kamus komprehensif bahasa Melayu". Penerbit Fajar Bakti, 2005.

[11] Le, T. A., Moeljadi, D., Miura, Y., & Ohkuma, T. "Sentiment analysis for low resource languages: A study on informal Indonesian tweets". In Proceedings of the 12th Workshop on Asian Language Resources (ALR12) (pp. 123-131), December 2016.

[12] Jamali, N. "Fenomena Penggunaan Bahasa Slanga dalam Kalangan Remaja Felda di Gugusan Felda Taib Andak: Suatu Tinjauan Sosiolinguistik". Jurnal Wacana Sarjana, 2(3), 1-1, 2018.

[13] Derczynski, L., Ritter, A., Clark, S., & Bontcheva, K. "Twitter part-of-speech tagging for all: Overcoming sparse and noisy data". In Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013 (pp. 198-206), 2013.

[14] Nakagawa, T., Kudoh, T. & Matsumoto, Y. "Unknown word Guessing and Part-of-Speech Tagging Using Support Vector" Mac, Hines. In Proceedings of the Sixth Natural Language Processing Pacific Rim Symposium, pp. 325—331, 2001.

[15] Zhu, B., Tokuno, J., & Nakagawa, M. "Segmentation of online handwritten Japanese text using SVM for improving text recognition". In International Workshop on Document Analysis Systems (pp. 208-219). Springer, Berlin, Heidelberg, February 2006.

[16] Giménez, J., & Marquez, L. "Fast and accurate part of speech tagging: The SVM approach revisited". Recent Advances in Natural Language Processing III, 153-162, 2004.

[17] Lee, Y. K., & Ng, H. T. "An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation". In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 41-48). Association for Computational Linguistics, July 2002.

[18] Schmid, H. "Part-of-speech tagging with neural networks". In Proceedings of the 15th conference on Computational Linguistics-Volume 1 (pp. 172-176). Association for Computational Linguistics, August 1994.

[19] Abdulkareem, M., & Tiun, S. "COMPARATIVE ANALYSIS OF ML POS ON ARABIC TWEETS". Journal of Theoretical & Applied Information Technology, 95(2), 2017.

[20] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. "Scikit-learn: Machine learning in Python". The Journal of Machine Learning Research, 12, 2825-2830, 2011.

[21] Bird, S., Klein, E., & Loper, E. "Natural language processing with Python: analyzing text with the natural language toolkit". "O'Reilly Media, Inc.", 2009.

[22] Kulkarni, A., & Shivananda, A. "Natural language processing recipes". Apress, 2019.

[23] Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., . . . Smith, N. A. "Part of speech tagging for twitter: Annotation, features, and experiments". Paper presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, 2011.

[24] Nooralahzadeh, F., Brun, C., & Roux, C. "Part of speech tagging for french social media data". In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers (pp. 1764-1772), August 2014.

[25] Màrquez, L., & Rodríguez, H. "Part-of-speech tagging using decision trees". In European Conference on Machine Learning (pp. 25-36). Springer, Berlin, Heidelberg, April 1998.

[26] Márquez, L. "Part-of-speech Tagging: A Machine Learning Approach based on Decision Trees". Universitat Politècnica de Catalunya, 1999.

Deep Learning Framework for Locating Physical Internet Hubs using Latitude and Longitude Classification

El-Sayed Orabi Helmi[1], Osama Emam[2], Mohamed Abdel-Salam[3]
Dept. Business Information Systems, Faculty of Commerce and BA, Helwan University, Cairo, Egypt[1, 3]
Dept. of Computer Science, Faculty of Computing and AI, Helwan University, Cairo, Egypt[2]

***Abstract*—This article proposes framework for determining the optimal or near optimal locations of physical internet hubs using data mining and deep learning algorithms. The framework extracts latitude and longitude coordinates from various data types as data acquisition phase. These coordinates has been extracted from RIFD, online maps, GPS, and GSM data. These coordinates has been class labeled according to decision maker's preferences using k-mean, density based algorithm (DB Scan and hierarchical clustering analysis algorithms. The proposed algorithm uses haversine distance matrix to calculate the distance between each coordinates rather than the Euclidian distance matrix. The haversine matrix provides more accurate distance surface of a sphere. The framework uses the class labeled data after the clustering phase as input for the classification phase. The classification has been performed using decision tree, random forest, Bayesian, gradient decent, neural network, convolutional neural network and recurrent neural network. The classified coordinates has been evaluated for each algorithms. It has been found that CNN, RNN outperformed the other classification algorithms with accuracy 97.6% and 97.9% respectively.**

***Keywords*—Physical internet hubs (π hubs); deep learning; convolutional neural network (CNN); recurrent neural network (RNN); latitude and Longitude classification**

I. Introduction

Choosing the location of storage warehouses is one of the most important steps facing supply chain officials, if not the most important step. This step has gained importance because of its impact on the entire supply chain. It also affects transportation operations and determines the speed of response to customer requests. Therefore, it was necessary for many researchers and specialists in supply chains to use all available means and techniques to determine the best location for these warehouses. With the rapid development of communication technologies and their overlap with the Internet of Things, large amounts of data became available for analysis and to give accurate mathematical alternatives to solve the problem of choosing the locations of those repositories. There are devices to track the movements of cars and their various effects along the supply chain. These devices also show the time of movement and waiting for these cars in an accurate and round-the-clock manner. There were many ways to analyze this data to reach the optimal location for these repositories. Classification processes are one of the most used methods to reach the most accurate solutions. The researchers also used other methods, such as relying on the global positioning system technology to track trucks in various places. From this standpoint, the research team decided to use the available data from that technology, taking into account the waiting hours, to determine the best places for these warehouses. But before proceeding in detail in this research, some concepts related to the topic of research must be presented, including first the different data collection techniques and the reason for choosing some of them over the other. Second, the methods of data analysis and the algorithms used to analysis that data. Third, the research team will address the new phenomenon in the field of supply chains, namely the physical Internet (PI), and the extent of its expected impact in the near future, in context of what has been published on this subject in some international scientific journals. Fourth, a detailed overview of deep learning techniques in the context of the research topic will be illustrated.

First vehicles tracking [1] Global Positioning System, The Global Positioning System (GPS), formerly known as Navstar GPS, is a satellite-based radio navigation system owned by the US government and maintained by the US Space Force. The GPS does not require the user to send any data, and it works independently of telephonic or Internet reception, however both technologies can improve the accuracy of GPS positioning data. Military, civic, and commercial users all across the world rely on the GPS for crucial positioning. The US government designed, maintains, and administers the system, which is publicly accessible to anybody with a GPS device. Based on data received from several GPS satellites, the GPS receiver estimates its own four-dimensional position in space-time. Each satellite keeps a precise record of its position and time, which it sends to the receiver which in our case is attached to trucks.

On other hand, Radio Frequency Identification (RFID) is a passive wireless technology that allows an item or person to be tracked or identified. Tags and readers are the two main components of the system. Radio Frequency Identification (RFID) relies on a small electrical device, generally a microchip, to store data. These devices are usually quite small, about the size of a grain of rice, and can store a lot of information. Some contain a stored power source or batteries, even if they don't always emit electricity. The scanners that are used to read these devices can also offer enough power to read the microchip. The technology has a variety of applications, but it is most typically used to track objects.

Global System for Mobile Communication (GSM) based tracking system is presented that uses a mobile phone text message system to keep track of a vehicle's location and speed. The technology may send text notifications for speed and location in real time. One of the disadvantages of this technology is that it bears an additional cost and weak or no signals in some places during the supply chain.

On other hand, researcher used computer vision approach for latitude and Longitude coordinates extraction [2]. A lot of data is required for computer vision. It repeats data analysis until it detects distinctions and, eventually, recognizes images. To teach a computer to recognize automotive tires, for example, it must be fed a large number of tire photos and tire-related materials in order for it to understand the differences and recognize a tire, particularly one with no faults. Computer vision enables procedures to be automated, saving time and reducing the pressure on a limited labor supply. Computer vision has been used to unload inventory trailers with increased efficiency. Vehicle departure times must be carefully synchronized with the loading periods of other trucks, trains, or planes.

The motive that motivated the research team to conduct such a study is the lack of a single framework that can deal with different types of data, whether classified or unclassified data. The team believes that by creating a single framework that integrates many data sources and provides different mining and machine learning algorithms and techniques, it can provide an addition in supply chain and physical internet operations, especially in the current and future period.

This article is divided into several parts. The first part provides solid overview of the main concepts of physical internet and deep learning techniques. It also, provides a brief summary of some related researches in the logistics industry. The second part discusses in detail the proposed framework. The last part demonstrates the experiments and results.

II. Theoretical Background

The following section discusses briefly the main concept of physical internet and deep learning. Some related studies will be discussed in section.

A. *Physical Internet*

The physical internet (PI) was developed in response to the inefficiencies and unsustainable nature of current logistics and supply chain management strategies. It was proposed to address issues that make current logistics practices unsustainable, such as limited space utilization for road, rail, sea, and air transportation, empty travel being the norm rather than the exception; poor working conditions for truck drivers, products sitting idle, inefficiency of product distribution, inefficient use of production and storage facilities, mediocre coordination within distribution networks, and high inefficiency of multimodal transportation [3].

The primary goal of the PI is to change "the way physical goods are handled, moved, stored, realized, supplied, and used, with the goal of improving global logistics efficiency and sustainability." The PI intends to coordinate physical commodities transit in the same way as data packets are carried via the digital Internet. The movement of commodities will be optimized in terms of cost, speed, efficiency, and sustainability by pooling resources such as vehicles and data and constructing transit centers that enable smooth interoperability. The PI establishes common and generally agreed-upon standards and protocols to promote horizontal and vertical cooperation amongst businesses in order to achieve this optimization.

The PI does not directly manipulate physical items, but rather manipulates and maintains the transportation containers that house them, just as digital Internet packets store embedded data. Many factors must be coordinated for the PI to become fully functional, including tangible items such as PI modular containers or PI transit centers, as well as more abstract notions such as legislation and business models. Previous research, which concentrated on performing simulations with a small number of participants in specific industries, has demonstrated that the application of the PI, even if it is partial, can result in significant advantages [4].

As shown in Fig. 1, according to ALICE [5] The Physical Internet contemplates the transformation of Logistics Nodes into Physical Internet nodes with standardized service definition and operations. Services at PI nodes are visible and digitally accessible to businesses, and they cover planning, booking/transactions, execution, and information exchange. The Physical Internet is based on a comprehensive and systemic consolidation of flow and network of networks principles. The Physical Internet promotes full consolidation of logistics flows from independent shippers in logistics networks (e.g., extended pooling). The Physical Internet promotes pooling resources and assets in open, connected, and shared networks (i.e. connecting existing (business) networks, capabilities, and resources) so that network users and partners can use them easily. It is predicted that by pooling demand and resources to meet that demand, resource utilization will be more efficient. Transport, storage, and physical handling procedures of load units such as containers, swap-bodies, pallets, boxes, and so on are included in the Physical Internet, as is any other resource required for a freight transport and logistics operation [5].

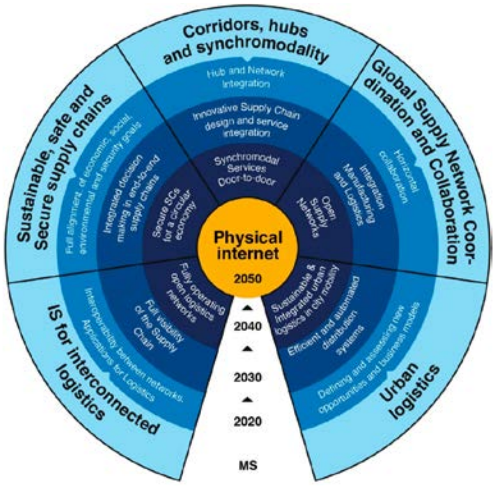

Corridors, hubs and synchromodality
Hub and Network Integration
Innovative Supply Chain design and service integration
Synchromodal Services Door-to-door
Global Supply Network Coordination and Collaboration
Horizontal collaboration
Integration Manufacturing and Logistics
Open Supply Networks
Sustainable, safe and Secure supply chains
Full alignment of economic, social, environmental and security goals
Integrated decision making in end-to-end supply chains
Secure SCs for a circular economy
Physical internet
2050
2040
2030
2020
MS
IS for interconnected logistics
Interoperability between networks. Applications for Logistics
Full visibility of the Supply Chain
Fully operating open logistics networks
Urban logistics
Defining and assessing new opportunities and business models
Efficient and automated distribution systems
Sustainable & integrated urban logistics in city mobility

Fig. 1. Physical Internet according to the European Technology Platform ALICE [5].

B. Deep Learning

Deep learning is a machine learning technique that trains computers to do what people do instinctively. Deep learning is a major technology underpinning a lot of applications such as automated driving, medical research, voice control, text, inventory management and other critical fields [6]. Deep learning has received a lot of attention recently, and for good reason. It is attaining results that were previously unthinkable. Deep learning models can attain cutting-edge accuracy, sometimes outperforming humans. Models are trained utilizing a huge quantity of labeled data and multi-layered neural network architectures [7] Fig. 2 demonstrates architecture of deep learning network layers. Deep learning has several known algorithms such as convolution neural network (CNN), recurrent neural network (RNN), long short term memory networks (LSTM) and generative adversarial networks (GANs). CNN and RNN are the main concern in this research. CNNs are multilayer perceptron that have been regularized. Multilayer perceptron are typically completely connected networks, in which each neuron in one layer is linked to all neurons in the following layer. Because of their complete connectedness, these networks are subject to data over fitting. Regularization, or preventing over fitting, is commonly accomplished by punishing parameters during training (such as weight decay) or reducing connectivity (skipped connections, dropout, etc.) CNNs use the hierarchical pattern in data to assemble patterns of increasing complexity utilizing smaller and simpler patterns imprinted in their filters. As a result, CNNs are at the lowest end of the connectivity and complexity spectrum [8]. RNNs are a form of neural network capable of modeling sequence data. RNNs, which are generated from feed forward networks, behave similarly to human brains. Simply said, recurrent neural networks are capable of anticipating sequential data in ways that other algorithms are not. RNNs feature a Memory that retains all calculation information. It uses the same parameters for each input because doing the same task on all inputs or hidden layers yields the same result.

Deep learning has several known algorithms such as convolution neural network (CNN), recurrent neural network (RNN), long short term memory networks (LSTM) and generative adversarial networks (GANs). CNN and RNN are the main concern in this research. CNNs are multilayer perceptron that have been regularized. Multilayer perceptron are typically completely connected networks, in which each neuron in one layer is linked to all neurons in the following layer. Because of their complete connectedness, these networks are subject to data over fitting. Regularization, or preventing over fitting, is commonly accomplished by punishing parameters during training (such as weight decay) or reducing connectivity (skipped connections, dropout, etc.) CNNs use the hierarchical pattern in data to assemble patterns of increasing complexity utilizing smaller and simpler patterns imprinted in their filters. As a result, CNNs are at the lowest end of the connectivity and complexity spectrum [8]. Fig. 3 illustrates CNN network architecture.

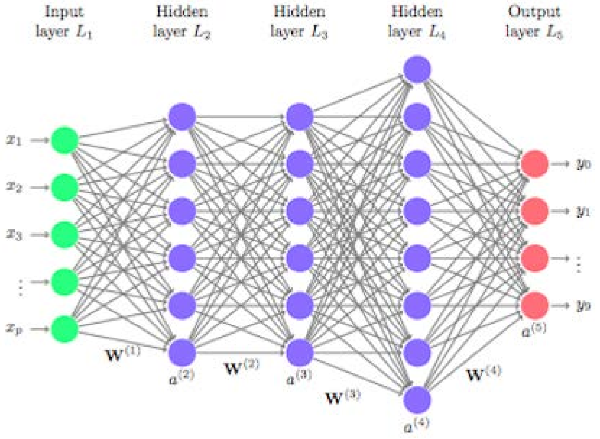


Fig. 2. Deep Learning Network Architecture [7].

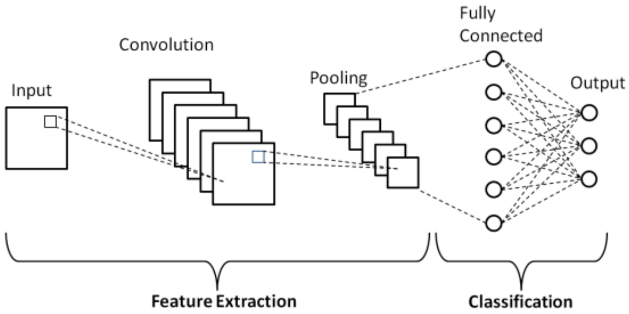


Fig. 3. Convolution Network Architecture [8].

RNNs are a form of neural network capable of modeling sequence data. RNNs, which are generated from feed forward networks, behave similarly to human brains. Simply said, recurrent neural networks are capable of anticipating sequential data in ways that other algorithms are not. RNNs feature a Memory that retains all calculation information. It uses the same parameters for each input because doing the same task on all inputs or hidden layers yields the same result.

C. Related Work

Several studies have been published to determine the best locations for supply chain warehouses using a number of different methods. Some of these researches used mathematical models to determine the locations and others used data mining techniques. In this section, the results of some of these studies, pointing out their advantages and disadvantages will be reviewed.

In [9], the researchers proposed an integrated Multi-Objective Hybrid Harmony Search-Simulated Annealing (MOHS-SA) algorithm to find the trade-off between the total cost of operating facilities, and the cost of CO2 emissions. The research presented a number of alternatives to warehouse locations, but it lacked any possibility of self-learning or determining the optimal location for these warehouses.

On other hand, in [10] the researchers examined a case study at a company that works in the Fast Moving Consumer Goods (FMCG) area and was undertaking a market test for its new alternative tobacco products using the B2C distribution system. The goal of this research is to identify both the number and location of warehouses that provide the lowest total logistics costs for the B2C process, while the company is still employing company-owned B2B warehouses. The number and location of these warehouses are determined using Agglomerative Hierarchical Clustering with Evolutionary Solver, where clustering is based on the shortest distance. One of the important effects that resulted from this research is that he was able to determine the number of warehouses based on the minimum costs, whether operating or transportation costs, in line with customer requests. The weak point of this research is that it locates warehouses based on cost only, without looking at other dimensions such as the environmental dimension, increased demand, or the introduction of other types of products. On other words this research is not the optimal solution for general case study.

In 2020 [11] proposed a mathematical model to allocate the warehouse centers in Far East of Russia based on population, trade turnover and volume of goods manufacturing. This research divided the warehouses to three categories federal, regional, and local centers. These logistics centers have been located in different cities regardless of the exact geographic information of these centers.

Using Lingo solver software [12] the researchers proposed a solution to locate warehouse based on linear programming. The decision in this research has been taken on cost of moving material, frequency trips of material, and the distance of location. This research suffered from the lack of the possibility of self-learning, as it relied entirely on linear mathematical equations without exposure to the non-linear nature of this problem.

In [13], the researchers proposed a solution to determine the optimal location of supply chain distribution centers using k-near neighbor clustering algorithm. They used spatial dataset in their case study. They have demonstrated the ability of this algorithm to divide geographical areas and locate warehouses. The defect of this algorithm was the necessity to determine the number of warehouses before starting the process of clustering the spatial data.

The authors of this paper [14] proposed a two-step clustering approach. In the first stage, they extract the frequent GPS trajectory halt locations. The second stage is to use a density-based clustering method to determine the nearest locations. Because it starts clustering from extracted points, the suggested approach saves time.

In [15], the authors proposed a clustering model using automatic identification system and DBSCAN algorithm to spot the location of ships using the navigation GPS latitude and longitude. The residence point of each ship is identified according to the ship speed and course change. This research uses only GPS data and lacks of self-learning possibility.

III. Proposed Framework

In this part, the proposed framework for solving the problem of locating supply chain repositories in context of the physical internet phenomenon will be presented. The proposed framework consists of four phases, data acquisition, data labeling, classification and testing. During data acquisition and selection phase, data from many sources is combined into a single data repository, resulting in a target dataset containing intriguing variables or data samples for discovery.

Due to the lack of classified data in many cases in actual supply chain applications and to maximize the effectiveness of the proposed framework, the research team performed a preliminary clustering of the collected data as data labeling phase. The clustering stage will be done using three known algorithms K-mean, density based algorithm (DB Scan) and hierarchical cluster analysis (HCA) to satisfy most of the decision-makers requirements. For example, when using k-mean algorithm, the decision maker can determine the number of PI hubs according to the available resources and future plans of the organization such equipment, transportation and workers. Decision makers can use the DB Scan algorithm to cluster the data without specifying a specific number of Hubs, due to the nature of the work of this algorithm. It divides the data into any number of groups so that the differences between the points of the same group are reduced to a minimum. If it is expected that there will be groups with arbitrary shapes, the decision maker can use DB Scan algorithm. The decision makers can use HCA to cluster the latitude and longitude coordinates according to regions or cities. The most significant aspect in hierarchical clustering is the linkage mechanism, which determines how the distances between clusters will be calculated. It has a significant impact on not just the clustering quality but also the algorithm's efficiency. The k-centroid link method has been chosen in our implementation. The output of this phase is labeled training dataset.

The clustering distance matrix used in the proposed framework is haversine formula. The haversine formula is a very accurate method of estimating distances between two places on the surface of a sphere given the two points' latitude and longitude. The haversine formula is a re-formulation of the spherical law of cosines; however the haversine formulation is more useful for tiny angles and distances. The central angle (θ) between any two points on a sphere is:

$$\theta = \frac{d}{r}$$

where:

d is the distance between the two points along a great circle of the sphere.

r is the radius of the sphere.

The haversine formula allows the haversine of θ (that is, hav(θ)) to be computed directly from the latitude (represented by φ) and longitude (represented by λ) of the two points:

$$hav(\theta) = hav(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\, hav(\lambda_2 - \lambda_1)$$

where:

φ1, φ2 are the latitude of point 1 and latitude of point 2,

λ1, λ2 are the longitude of point 1 and longitude of point 2.

Finally, the haversine function hav(θ), applied above to both the central angle θ and the differences in latitude and longitude, is

$$hav(\theta) = sin^2\left(\frac{\theta}{2}\right) = \frac{1 - \cos(\theta)}{2}$$

In the third phase, the proposed framework presents the possibility of using a number of well-known classification algorithms to classify data, such as: decision tree, Naïve Bayes, random forest, gradient descent, k-nearest neighbors, support vector machine (SVM), CNN, and RNN. Although some of these algorithms are lazy learning algorithms, they classify data accurately in many cases. The proposed framework uses CNN and RNN deep learning algorithms to add self-learning nature.

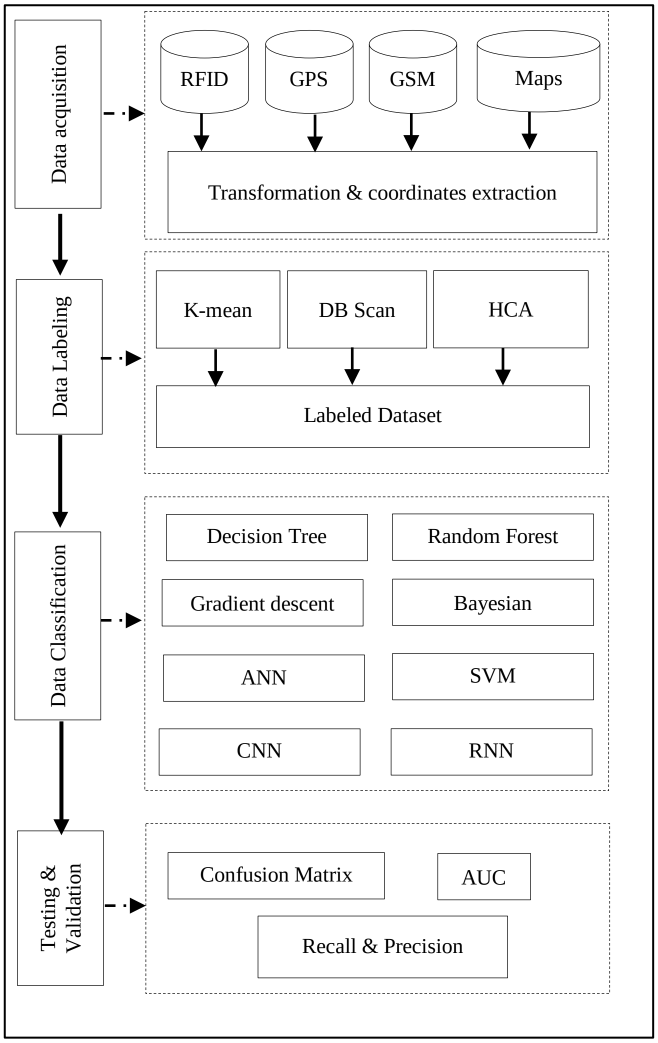


Fig. 4. Physical Internet Hubs Locating using Latitude and Longitude Classification.

Testing and validation is the fourth phase. At this phase, the results of all classification algorithms that have been used in the proposed framework are evaluated to find out the optimal outcomes within standardized testing criteria. In this phase the area under curve (AUC), confusion matrix and precision have been used to calculate the framework accuracy.

As in Fig. 4 the framework extracts latitude and longitude coordinates from four major resources (RFID, GPS, GSM, and maps) at data acquisition phase the framework has been developed to handle all of these data types. It is not necessary to use all data types in real case scenarios; any user can customize this step according to his application.

IV. Experiments and Results

This section discusses in detail the performed experiments. All experiments were carried out using Spain's coordinate's dataset on the Kaggle website [16]. The frame work has been implemented with python3 environment. Table I illustrates the Spain's coordinate's dataset main features.

All null features (city, district, & region) and missing instances have been dropped as preprocessing phase. The reset 975000 instances will be divided as 70% for training and 30% for testing and validation for the classification purposes after the preparation phase.

The dataset was divided into 15 class labeled groups using k-mean algorithm as data preparation phase. The number of groups was chosen to reduce distances to a minimum in order to reduce the time and cost of transporting goods between points which achieves sustainability. Then this labeled dataset has been fed to the proposed framework for classification as 3rd phase. The architecture implementation of CNN algorithms was as the following: 1 Conv. layer, 5 dense layers and the used activation function was Relu. The RNN implementation also as the following; 1 RNN layer, 4 dense function using tanh activation function. Table II shows the evaluating comparison of the used classification algorithms.

As shown in Table II, the Bayesian algorithm classification accuracy was 86.2%. The Random forest algorithm achieved 90.2% accuracy. CNN and RNN algorithms made the classification with 97.6%, and 97.9% respectively. On other hand SVM classified the dataset with accuracy 95.1%. The previous results have been calculated with confidence level 95%. The results showed also, that RNN outperformed the classification rather than other algorithms.

TABLE I. Spain Open Address Dataset Features

#	**Features**	**Unique Instances**
1	Latitude	977080
2	Longitude	977070
3	Street name	470045
4	City	Null
5	District	Null
6	Region	Null
7	Postal code	976000

TABLE II. THE PROPOSED FRAMEWORK ACCURACY COMPARISON

	AUC	**CA**	**Precision**	**Recall**
Bayesian	0.8956	0.862	0.885	0.890
Random Forest	0.9145	0.902	0.903	0.912
CNN	0.9806	0.976	0.971	0.978
RNN	0.9812	0.979	0.982	0.989
SVM	0.9342	0.928	0.913	0.922
ANN	0.9564	0.951	0.934	0.946
Gradient descent	0.9012	0.899	0.891	0.898
Decision tree	0.9102	0.909	0.906	0.904

V. CONCLUSION

The proposed framework consists of four major phases. The first phase is data acquisition. In this phase the latitude and longitude data extracted from different data sources such as: RFID, GSM, GPS, and maps. This data has been class label using one of three clustering algorithms (k-mean, DB scan, & HCA) as second phase. The third phase is classification phase. In this phase, the decision makers can choose from one or more classification algorithms such as: Bayesian, decision tree, random forest, SVM, ANN, CNN and RNN. The performed experiments showed that RNN classified the data better than other algorithms with accuracy 97.9%. According testing results, the research team believes that it is preferable to locate the physical internet hubs by one of deep learning algorithms especially CNN or RNN rather than other classification techniques. This study can be extended in the future by adding new deep learning algorithms or using different activation functions.

REFERENCES

[1] Sumit S. Dukare, Dattatray A. Patil, Kantilal P. Rane, Vehicle Tracking, Monitoring and Alerting System: A Review, International Journal of Computer Applications, Vol. 119, pp. 39-40, 2015.

[2] Bo Yang, Mingyue Tang, Shaohui Chen, Gang Wang, Yan Tan & Bijun Li., A vehicle tracking algorithm combining detector and tracker, Journal on Image and Video Processing, Vol. 17, pp. 10-17, 2020.

[3] Horst Treiblmaier, Kristijan Mirkovski, Paul Benjamin Lowry, Zach G. Zacharia., the physical internet as a new supply chain paradigm: a systematic literature review and a comprehensive framework, The International Journal of Logistics Management, Vol. 31, pp. 240-280, 2020.

[4] Sarraj, R., Ballot, E., Pan, S., Hakimi, D. and Montreuil, B, Interconnected logistic networks and protocols: simulation-based efficiency assessment, International Journal of Production Research, Vol. 52 , pp. 3185-3208, 2014.

[5] Klumpp, Matthias, Automation and artificial intelligence in business logistics systems: human reactions and collaboration requirements, International Journal of Logistics, Vol. 21, pp. 224- 242, 2017.

[6] Vega-Márquez B,Nepomuceno-Chamorro I,Jurado-Campos N and Rubio-Escudero C , Deep Learning Techniques to Improve the Performance of Olive Oil Classification, Frontiers Chemistry, Vol. 7, pp. 1-10, 2020.

[7] Vankara, J., Krishna, M.M., Dasari, S, Classification of Brain Tumors Using Deep Learning-Based Neural Networks, , Smart Technologies in Data Science and Communication, Singapore, Springer , pp. 33-40, 2021.

[8] Avilov, Oleksii; Rimbert, Sebastien; Popov, Anton; Bougrain, Laurent. Montreal , Deep Learning Techniques to Improve Intraoperative Awareness Detection from Electroencephalographic Signal,2nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 142–145, 2020.

[9] F Misni, L S Lee, N I Jaini, Multi-objective hybrid harmony search-simulated annealing for location-inventory-routing problem in supply chain network design of reverse logistics with CO2 emission, Journal of Physics: Conference Series, pp. 1-16, 2021

[10] Nyoman Sutapa, Magdalena Wullur, Tania Nano Cahyono, Determining the Number and Location of Warehouses to Minimize Logistics Costs of Business to Consumer (B2C) Distribution, SHS Web of Conferences, pp. 1-8, 2020.

[11] A Bardal , M Sigitova, Localization of Transport and Logistics Centers in the Region, IOP Conf. Series: Materials Science and Engineering, pp. 1-6, 2020.

[12] Ade Irman, Y Muharni, Andri Yusuf. Bali, Design of warehouse model with dedicated policy to minimize total travel costs: a case study in a construction workshop: International Conference on Advanced Mechanical and Industrial engineering, Indonesia pp. 1-7, 2020.

[13] Sumit S. Dukare, Dattatray A. Patil, Kantilal P. Rane., Vehicle Tracking, Monitoring and Alerting System: A Review, International Journal of Computer Applications, Vol. 119, pp. 29-40, 2015.

[14] Bo Yang, Mingyue Tang, Shaohui Chen, Gang Wang, Yan Tan & Bijun Li, A vehicle tracking algorithm combining detector and tracker.. s.l. : J Image Video Proc., Vol. 17, pp. 10-20, 2020.

[15] Horst Treiblmaier, Kristijan Mirkovski, Paul Benjamin Lowry, Zach G. Zacharia., the physical internet as a new supply chain paradigm: a systematic literature review and a comprehensive framework, The International Journal of Logistics Management, Vol. 31, pp. 240-280, 2020.

[16] Open address Europe. www.Kaggle.com. [Online] May 10, 2017. [Cited: J https://www.kaggle.com/datasets/openaddresses/openaddresses-europe?select=spain.csv, last access: 5 june 2022..

Development of Adaptive Line Tracking Breakpoint Detection Algorithm for Room Sensing using LiDAR Sensor

Deddy El Amin, Karlisa Priandana, Medria Kusuma Dewi Hardhienata
Computer Science Department, Bogor Agricultural University (IPB University), Bogor, Indonesia

***Abstract*—This research focuses on the use of Light Detection and Ranging (LiDAR) sensors for robot localization. One of the most essential algorithms in LiDAR localization is the breakpoint detector algorithm which is used to determine the corner of the room. The previously developed breakpoint detection methods have weaknesses, such as the Adaptive Breakpoint Detector (ABD), could generate dynamic threshold values. The ABD results, on the other hand, still require Line Extraction to obtain the corner breakpoint. Line Extraction method, e.g. Iterative End Point Fit (IEPF), is used to categorize data, resulting in the generation of a line pattern as an interpretation of a wall. The computational method for obtaining the corner breakpoint becomes longer as the line is extracted. To address this issue, our algorithm proposes a new threshold area in the form of an ellipse with the threshold value parameter obtained from previously identified room size and sensor characteristics. As a result the corner breakpoint detection becomes more adaptive. The goal of this research is to create an Adaptive Line Tracking Breakpoint Detector (ALTBD) approach that will reduce the computing time required to detect corner breakpoints. Furthermore, the Line Extraction method required for corner breakpoint detection is modified in the ALTBD. To distinguish between the edge of the wall and the corner of the room, the boundary value is increased. The ALTBD method was tested in a simulation arena comprised of multiple rooms and halls. According to the results, the ALTBD computation time is faster in detecting corner breakpoints than the ABD IEPF method, also the accuracy for determining the position of the robot was improved.**

Keywords—LiDAR; breakpoint detector; robot localization; corner detection; line segmentation

I. Introduction

Robot localization which is a method to acquire information about a robot's direction and position in its working environment is required by every autonomous robot in order to move and accomplish its task. Three types of robot localization are global localization, predictive localization, and local localization [1]. When the robot is outside, global localization works relatively well [2]. Whenever the robot is indoors, it uses probabilistic localization, local localization, or a combination of the two [3]. Our research leads to probabilistic localization, which determines the robot's position and orientation based on sensor measurement data combined with prior knowledge in the form of a map of the arena in the room.

Several papers have been written about the sensors used in localization and probabilistic localization methods, including encoder sensors [4][5], magnetic compass [6], magnetic anomalies map [7], low-cost gyro [8], ultrasonic sensors [9] [10], RFID [11] to the Light Distance and Ranging (LiDAR) sensor. Sensors such as encoder disc sensors, magnetic compasses, low-cost gyros, and ultrasonic sensors require significant movement before the robot can identify its position. If time is of the essence in accomplishing the objective of the robot, the sensor is not the first option to be installed in the robot. As a result, this research utilizes a LiDAR sensor capable of identifying the position and surroundings without requiring the robot to move, improving the robot's movement more effective.

Previous research has widely proposed the LiDAR sensor since it has excellent spatial accuracy, fast data renewal, and does not rely on object illumination or reflectance [12]. This research focuses on determining the form of the room using the LiDAR sensor measurement results. The issue is determining how to identify the geometry of the space based on its corner position. Breakpoint detection is one way for determining the corner of a room. This method detects the room's corner points and the edge of the wall by verifying the discontinuity between the two points scanned consecutively by the LiDAR sensor [13].

Breakpoint detection method that combines Successive Edge Following (SEF), Line Tracking (LT), and Iterative End Point Fit (IEPF) was developed by Siadat et al. [14]. However, all three methods have a weakness: finding a consistent threshold value to detect a breakpoint between two consecutive sensor scan points is challenging. The observed breakpoints can take the form of room corners or the edge of a wall. The breakpoint detection threshold value should be dynamic, based on the distance between the sensor and the object being scanned [13].

Several researchers have indicated determining dynamic threshold values. The threshold value in Lee et al. [15] research is based on the current and prior measurement distance values. The DIET method, which takes the measurement angle into account when establishing the threshold value, was developed by Dietmayer et al. [16]. The method by including the virtual wall angle parameter as an estimate of the detected wall slope angle was improved by Santos et al. [17]. The threshold value in their Adaptive Breakpoint Detector (ABD) method as well developed by Borges and Aldon [13]. They use auxiliary angles to determine the point of intersection and use it as a reference for the threshold value, and Certad et al. [18] also adjusted the threshold value in the ABD. They modify the measurement distance value in the ABD equation with the shortest distance value from either the current or previous distance. In his research on the ABD method, Su Young et al. [19] identified the lack of a line cluster perpendicular to the sensor as a problem. The challenge was then solved by combining the ABD and LT methods to create the Dual Breakpoint Detector (DBD), and Weerakoon et al. [20] also used ABD in their research. In order to increase the feature extraction rate value, Weerakoon et al. [20] research suggests taking the measurement error value and the longest distance value into consideration. The corner breakpoint was determined in three phases in this investigation. The first stage is ABD, which is used to segment the data depending on a threshold value. The following stage is Line Extraction with IEPF. Line Extraction is a method for organizing data based on line patterns that can be formed as an interpretation of a wall. The intersection of the two lines obtained by the IEPF two-segment linear regression is then used to locate the corner breakpoint in the third phase. Dingyao et al. [21] research's to determine the corner breakpoint included one more phase into four phases, i.e. point feature matching to compare angles and distances. The second, third and fourth stages of Corner Breakpoint Detection render the algorithm's computational process worthless.

To overcome these issues, The Adaptive Line Tracking Breakpoint Detector (ALTBD) method is developed in this research. The ALTBD method is a variation of the Line Tracking (LT) and Adaptive Breakpoint Detector (ABD) algorithms that introduces a new threshold area to improve corner breakpoint detection. By using established room sizes, this method also solves the challenge of finding the threshold. The breakpoint detection threshold value is made dynamic in this method by taking into consideration the difference between the sensor measurement distance and the projected virtual wall distance. Furthermore, the Line Extraction method required for corner breakpoint identification is improved in the ALTBD to reduce computational time. To distinguish between the edge of the wall and the corner of the room, the boundary value is increased.

This paper outline is as follows: The next section reviews some relevant works on LT and ABD method. Section III defines the proposed adaptive line tracking breakpoint detector method. Section IV provides the methodology that used to in this research, and Section V shows the results and analysis, whilst the conclusion is presented in the final section.

II. Related Work

In this section, several methods are discussed as the basis for the proposed method, namely Line Tracking (LT) and Adaptive Breakpoint Detector (ABD) method.

A. Line Tracking Method

Line tracking (LT) is a method of detecting breakpoints or segmenting line data that use linear regression and works in Cartesian coordinates. The following pseudo code [22] describes the working principle of this method:

1) Begin with two points of measurement, then draw a line between them.

2) Insert the next point to make a new line model.

3) Reprocess the line's parameters.

4) If the result of the generated line is satisfactory, continue (repeat to step 2).

5) If the result is not satisfactory, make the last point the line segment's end point, re-process the line's parameters, and make it as one line segment.

6) Go back to step 2 after completing the next two points.

The difficulty in defining the threshold value (D_{max}) to determine the breakpoint between the two measures is this LT's disadvantage [13]. Fig. 1 depicts two segments of data generated by LT when $d_5 > D_{max.}$

B. Adaptive Breakpoint Detector Method

The Adaptive Breakpoint Detector (ABD) method to address the issue of a fixed threshold value at LT was proposed by Borges and Aldon [13]. This method performs on the same principles as the LT method, which utilizes a threshold to detect breakpoints. The distinction is that there is no linear regression in this method, and the threshold value is made dynamic by utilizing the angle difference between the two measurement results. The ABD method is illustrated in Fig. 2 along with the parameters utilized. In ABD, breakpoints are detected by creating a border circle with a radius of D_{max} and a center at p_{n-1}. A breakpoint is detected if the current scan point (p_n) is found to be outside the circle. As a result, the previous scan point (p_{n-1}) is considered the end point of the current segment, whereas p_n is considered the beginning point of the following segment.

The adaptive breakpoint detector (ADB) method has the advantage of having a threshold value (D_{max}) that can change depending on the previously created distance (r_{n-1}) and a difference in the angle of measurement ($\Delta\phi$). This value is used as a comparison (border value) to distinguish between two segments of line data.

Equation (1) [13] describes the data terms stated as breakpoints:

if $\|p_n - p_{n-1}\| > D_{max}$ then k^b_n := TRUE and k^b_{n-1} := TRUE
(1)

Where, $\|p_n - p_{n-1}\|$ represents the distance between the p_n point and the p_{n-1} point. D_{max} is the breakpoint threshold whose value is determined by the sensor's distance to a measurable object (r_n). The current measured position is p_n, while the previous measured point is p_{n-1}. Fig. 2 depicts this ADB method. Detection is accomplished by drawing a dividing circle with the center point at p_{n-1} and a radius of D_{max}.

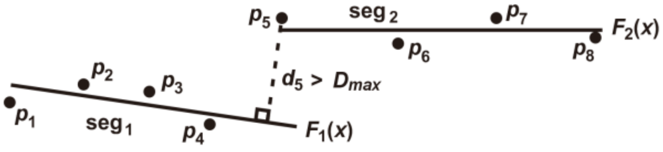


Fig. 1. Line Tracking Method [14].

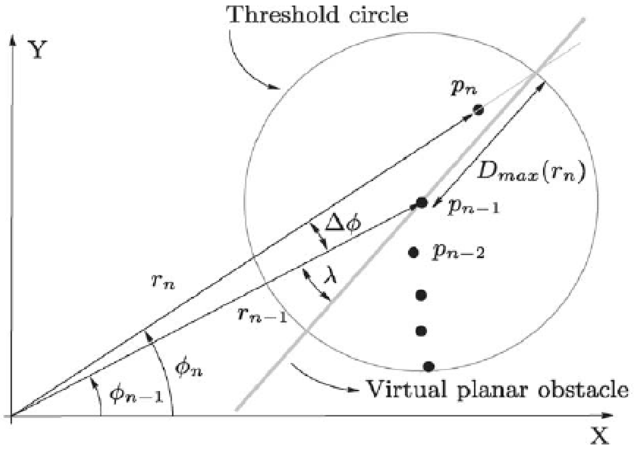


Fig. 2. Adaptive Breakpoint Detector Method [4].

If the current scan point (p_n) is found to be outside the circle, $p_{n\text{-}1}$ is considered the endpoint of the current segment and p_n is determined the start point of the following segment. Equation (2) [13] describes its mathematical function:

$$\|p^h{}_n - p_{n\text{-}1}\| = r_{n\text{-}1}.\ (\sin(\Delta\phi)/\sin(\lambda\text{-}\Delta\phi)) \qquad (2)$$

$\|p^h{}_n - p_{n\text{-}1}\|$ as the breakpoint testing threshold value Equations (3) and (4) are provided a real implementation of the threshold formula with [13]:

$$D_{max} = \|p^h{}_n - p_{n\text{-}1}\| + 3\sigma_r \qquad (3)$$

$$D_{max} = r_{n\text{-}1}.\ (\sin(\Delta\phi)/\sin(\lambda\text{-}\Delta\phi)) + 3\sigma_r \qquad (4)$$

Where, σ_r is the laser scanner manufacturer's parameter and λ the angle between the virtual line and the line connecting the object point to the laser scanner ($r_{n\text{-}1}$).

1) Line extraction: The ABD method generates breakpoints only at the ends of segments, which we refer to as terminal breakpoints in our research. As a result, another algorithm, Line Extraction, is required to reprocess existing data into new segments that can represent a line. As a result, the ABD computation time exceeds the LT in determining the angular breakpoint. Linear Regression [23], Split and Merge algorithm [24], Iteration End Point Fit [25], RANSAC algorithm [26], Hough Transform algorithm and Expectation-Maximization algorithm are some line extraction algorithms that can be used [27]. The Iteration End Point Fit (IEPF) method is discussed in this chapter as a comparison to the new method developed.

The IEPF algorithm [25] is a recursive Line Extraction method. Data checks are performed repeatedly for each segment base due to its recursive nature. The IEPF method is depicted in Fig. 3 by truncating one segment of data resulting from ABD into three data segments that form three lines. First, the initial data (p_1) and final data (p_{n}) are connected to form a straight line. The distance between the second data and the line is then calculated orthogonally. This is repeated for the third data point, and so on until all members of the segment have been checked. The maximum orthogonal distance ($d_i{}^m$) of the result will be compared to the limit value (d_{thd}). The limit value is the comparison value that was declared at the beginning. If the maximum value ($d_i{}^m$) is greater than the limit value, the segment will be halved at the maximum distance point. Fig. 9 illustrates how the first cut of a segment (Z_1 = { p_1,.., p_n}) into two segments (Z_2 = {p_1,..., p_i} and Z_3 = {p_i,..., p_n}) occurs at the p_i point. The initial data (p_1) and final data (p_i) of the Z_2 segment are then connected to form a straight line. The distance between all of the data points in the segment Z_2 is then calculated orthogonally to the line. The result's maximum orthogonal distance ($d_h{}^m$) will be recompromised by the limit value (d_{thd}). One line segment has been checked because the maximum orthogonal distance ($d_h{}^m$) is less than the limit value (d_{thd}). The segment is a Z_2 made up of p_1 to p_i. The Z_3 segment is treated similarly. However, the maximum orthogonal distance ($d_j{}^m$) in the Z_3 segment {(p_i,..., p_n)} is greater than the limit value (d_{thd}), so the segment must be cut in half with the intersection point at the p_j. The maximum orthogonal distance between the segments {p_i,..., p_j} and {p_j,..., p_n} is then calculated and compared to the limit value (d_{thd}). The data grouping on the Z_1 segment is complete because the maximum orthogonal distance of each segment is less than the limit value (d_{thd}). Through IEPF on the Z_1 segment, the final result is three line segments, namely S_1{p_1,..., p_i}, S_2{p_i,..., p_j}, and S_3{p_j,..., p_n}.

Weerakoon et al. [20] research's describes examples of ABD and Line Extraction implementation. The ABD threshold value equation was modified to Equation (5) in the research [20].

$$D_{max} = min\{r_n,r_{n+1}\}(1 - 3\sigma_r)(\sin(\Delta\phi)/\sin(\lambda\text{-}\Delta\phi))$$
$$+ 3\sigma_r.max\{r_n,r_{n+1}\} \qquad (5)$$

The ABD method has a disadvantage in that the angle value generating the threshold value is obtained based on the results of the experiment and requires Line Extraction to detect the corner breakpoint.

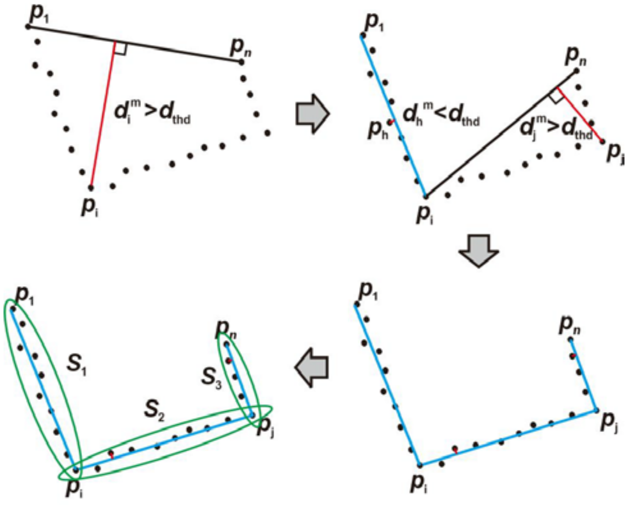


Fig. 3. IEPF Algorithm Working Principle.

III. THE PROPOSED ADAPTIVE LINE TRACKING BREAKPOINT DETECTOR METHOD

A. Propose a New Method (Adaptive Line Tracking Breakpoint Detector)

The Adaptive Line Tracking Breakpoint Detector (ALTBD) method we propose is a combination of the Line tracking (LT) and Adaptive Breakpoint Detector (ABD) methods. The existing breakpoint detector (ABD) obtains coverage in the form of a circle with a threshold value diameter, whereas ALTBD obtains coverage in the form of an elliptical area with an elliptical center point from the prediction point. The LT method calculation yielded the prediction point. Although elliptical shapes require more complicated equations, they depict a point spread rather than a circle. The threshold value on ABD is determined by the virtual angle derived from the experiment, whereas in ALTBD, the parameter producing an elliptical area, which is a minor value, is made adaptable by taking the measurement error of the LiDAR sensor into consideration.

The linear equation $F'_n(x)$ generated by LT is used to predict the next point of measurement, as shown in Fig. 4. The point's polar angle value (θ_{n+1}) is then used as an input to calculate the equation of the line $F_{n+1}(x)$. The predictive point p'_{n+1} is then obtained by intersecting the two equations of the line. The prediction point p'_{n+1} is defined as a point (0,0) or the center point of the elliptical area, with the x-axis representing as the major axis and the y-axis representing as the minor axis. The length of the major axis ($qmax_{n+1}$) is determined by the distance between the points p_n and the prediction line $F'_n(x)$ at the point of intersection. Minor axis length ($hmax_{n+1}$) is determined by calculating using orthogonal regression against the prediction line and the error values at the previous points. Fig. 4 and 5 depict the parameters used in the ALTBD method. The ALTBD algorithm's flow is described in detail:

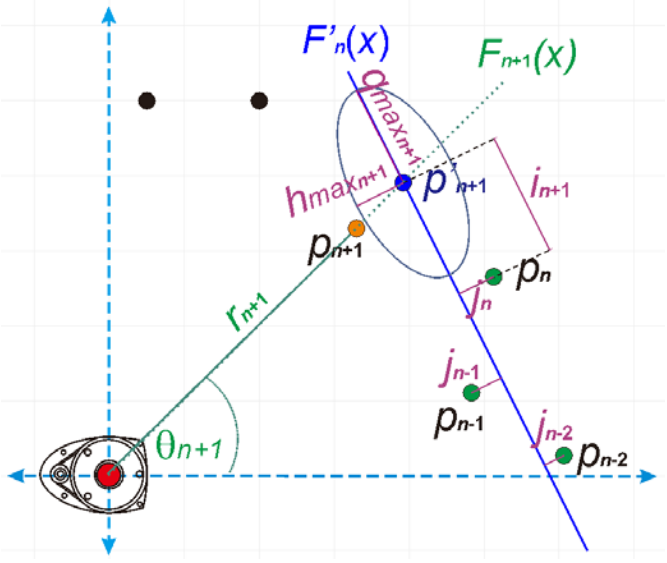


Fig. 4. ALTBD Method.

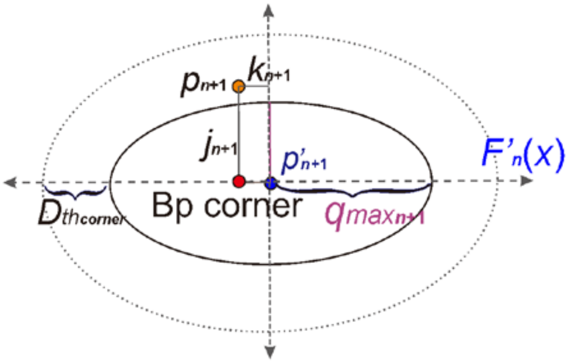


Fig. 5. Corner Detection ALTBD.

Algorithm for Detecting Breakpoints in Adaptive Line Tracking:

1) Determine the linear regression equation from three starting points (p_{n-2}, p_{n-1}, p_n) and transform it into a predictive line ($F'_n(x)$).
2) Determine the difference between the starting point (j_{n-2}, j_{n-1}, j_n) and the predicted line ($F'_n(x)$).
3) Using the equation $\boldsymbol{\varepsilon}_n = |j_n|/r_n$, calculate the percentage error ($\boldsymbol{\varepsilon}_{n-2}$, $\boldsymbol{\varepsilon}_{n-1}$, $\boldsymbol{\varepsilon}_n$) relative to the distance to the sensor (r_n).
4) Determine the average error percentage, $\sigma_n = (\Sigma^n_{i=n-2}\boldsymbol{\varepsilon}_i)/3$
5) Determine the minor axis limit, $hmax_{n+1} = \sigma_n + \sigma_{offset}$.
6) Verify out the next angle data (θ_{n+1}).
7) Create an equation using the angle gradient (θ_{n+1}) and the line ($F_{n+1}(x)$) that passes through the point (0,0).
8) Make the intersection point of the prediction line ($F'_n(x)$) and the scan line ($F_{n+1}(x)$) the prediction point (p'_{n+1}).
9) Calculate the distance between the prediction point (p'_{n+1}) and the last scan point (p_n) parallel to the prediction line ($F'_n(x)$), and set it as the major axis limit ($qmax_{n+1}$).
10) Read the next data set (p_{n+1}).
11) Transform the predicted point (p_{n+1}) to new coordinates, with the predicted point (p'_{n+1}) as the origin (0,0) and the predicted line ($F'_n(x)$) as the x-axis.
12) Calculate the new position of the point (p_{n+1}), and update x by k_{n+1} and y by j_{n+1},
    $k_{n+1} = \cos(\theta).(xp_{n+1} - xp'_{n+1}) + \sin(\theta).(yp_{n+1} - yp'_{n+1})$
    $j_{n+1} = -\sin(\theta).(xp_{n+1} - xp'_{n+1}) + \cos(\theta).(yp_{n+1} - yp'_{n+1})$
13) Enter the minor axis boundary ($hmax_{n+1}$), the major axis boundary ($qmax_{n+1}$), and the new measurement point position (k_{n+1}, j_{n+1}) into the ellipse equation,
    $e\_result = ((k^2_{n+1})/(qmax_{n+1})) + ((j^2_{n+1})/(hmax_{n+1}))$
14) If $e\_result > 1$, a breakpoint has been detected. Then, determine whether the position of the point (p_{n+1}) is within the tolerance range of the corner (Dth_{corner}).
15) If it is within the corner's tolerance limit, set the point (k_{n+1},0) as the corner breakpoint and transform it back to the original coordinates.
16) If it is not within the corner's tolerance range, use the following three measurement points (p_{n+2}, p_{n+1}, p_n) as the starting point and return to step 1.
17) If $e\_result \leq 1$, then calculate the difference between the measurement points and the prediction line (j_{n+1}).
18) Recalculate the percentage of error, $\boldsymbol{\varepsilon}_{n+1} = |j_{n+1}|/r_{n+1}$
19) Recalculate the minor axis limit, $hmax_{n+1} = \sigma_{n+1} + \sigma_{offset}$
20) Repeat step 6.

IV. Methodology

A. Materials and Tools

This research was performed out with the following hardware and software: The RpLidar A1 module, Beaglebone Black Wireless, Intel i5 2.5 GHz RAM 4GB Win32bit Laptop, and Jupyter Notebook.

B. Stages of Research

The seven research stages in the development of the Adaptive Line Tracking Breakpoint Detector (ALTBD) method are as follows: data acquisition, LiDAR data noise model design, LiDAR data generator module design with simulation, data processing using the ALTBD method, data processing using comparison methods (ABD Line Extraction and LT), testing the ALTBD algorithm, and analysis and evaluation. Fig. 6 depicts the flow of the research method.

Because the data utilized in the test will use simulation data from the LiDAR data generator module, the first stage of this research is LiDAR data acquisition, which is done in real time to obtain a noise sensor model. The purpose of this data acquisition is to determine the noise characteristics of the RpLidar data and to validate the noise characteristics of the sensor manual document. Observations were taken on a straight wall with no obstacles that was greater than 5 meters long.

1) LiDAR data acquisition: The next stage is LiDAR data acquisition, which is done to create a LiDAR sensor noise model. The distance between the LiDAR sensor and the wall was varied from 15 to 225 cm with 5 cm intervals to obtain appropriate data. For each distance variation, data were collected five times.

2) Design of noisi model for the RpLiDAR A1: The simulation module use the noise model to verify that the generated data is as close to the real LiDAR data as possible. Residual noise will be added at random and generated using a normal distribution. The model is built using a polynomial regression method on the outcomes of LiDAR data collecting. Linear or polynomial interpolation is used to approximate the values of the element variables.

3) Design of a LiDAR data generator simulation module: The arena simulation module is used to test the ALTBD method. A square-shaped arena with four rooms and multiple tunnels is used for simulation testing. That's the 2.4 m by 2.4 m arena for the Indonesian Fire Extinguisher Robot Contest (KRPAI 2019). In the KRPAI guide document [28], the position of the doors in rooms 3 and 4 distinguishes four distinct room layout combinations. This simulation module is made up of five blocks, as depicted in Fig. 7. This module takes as input the robot's position (x,y) and direction (azimuth), as well as the choice of space combinations 3 and 4. This module generates level 3 scan data with angle and distance attributes.

4) Testing: The tests were analyzed by comparing three Breakpoint Detector methods: the one we developed (ALTBD), the LT method, and the ABD IEPF method from the Weerakoon et al. [20] research. In this research, the computing time, percentage of distance inaccuracy, and position difference between the Breakpoint Pattern Recognition results will be sought from each method. Computational time is measured beginning with level 3 scan data, which is processed in each method to produce an output in the form of a corner breakpoint. The percentage of distance measurement error is calculated by subtracting the results of the Breakpoint Detector from the position of the reference breakpoint.

The difference is then compared to the sensor's distance from the reference breakpoint. But first, the Breakpoint Detector data must pass through the Breakpoint Extraction module to enable determining the reference breakpoint point easier. This research will investigate the position difference between the Breakpoint Pattern Recognition results and the actual global position. On the LiDAR data generator simulation module, the global position is the input sensor location in the arena. Fig. 8 depicts the test diagram.

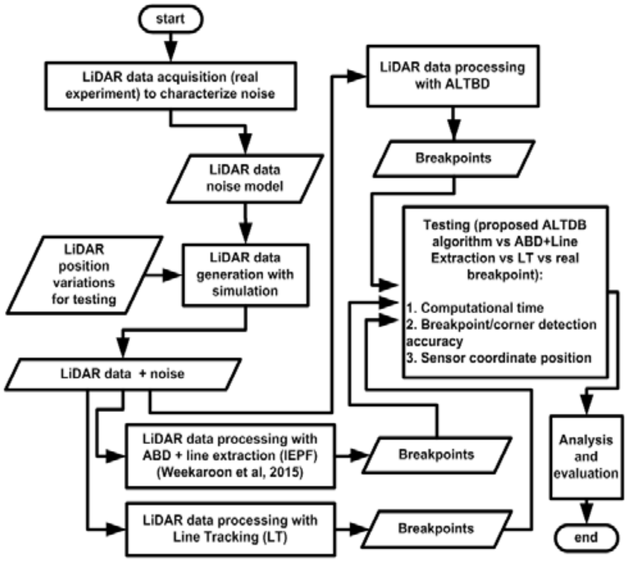

start
LiDAR data acquisition (real experiment) to characterize noise
LiDAR data noise model
LiDAR position variations for testing
LiDAR data generation with simulation
LiDAR data + noise
LiDAR data processing with ALTBD
Breakpoints
Testing (proposed ALTDB algorithm vs ABD+Line Extraction vs LT vs real breakpoint):
1. Computational time
2. Breakpoint/corner detection accuracy
3. Sensor coordinate position
LiDAR data processing with ABD + line extraction (IEPF) (Weekaroon et al, 2015)
Breakpoints
LiDAR data processing with Line Tracking (LT)
Breakpoints
Analysis and evaluation
end

Fig. 6. Research Stages.

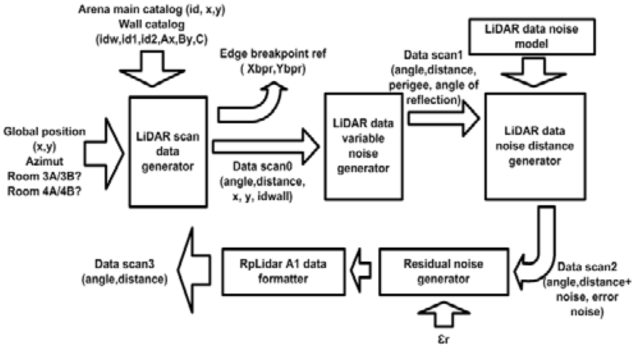


Fig. 7. Arena Simulation Module Block Diagram.

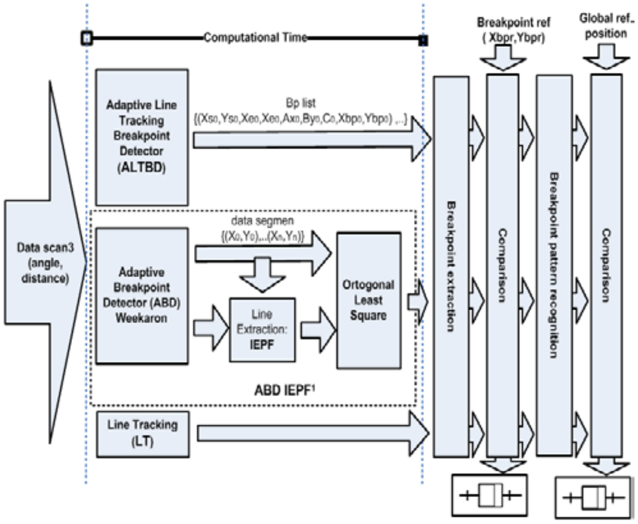

Computational Time
Breakpoint ref (Xbpr,Ybpr)
Global ref. position
Data scan3 (angle, distance)
Adaptive Line Tracking Breakpoint Detector (ALTBD)
Bp list {(Xs0,Ys0,Xe0,Xe0,Ax0,By0,C0,Xbp0,Ybp0) ,..}
data segmen {(X0,Y0),..(Xn,Yn)}
Adaptive Breakpoint Detector (ABD) Weekaron
Line Extraction: IEPF
Ortogonal Least Square
ABD IEPF¹
Line Tracking (LT)
Breakpoint extraction
Comparison
Breakpoint pattern recognition
Comparison

Fig. 8. ALTBD Test Method.

a) Breakpoint Extraction: This module sorts breakpoints, namely corner breakpoints, corner breakpoint terminal points, and other points that form only one side of the wall. This module also removes breakpoints that are overlapping. The breakpoint extraction module is required to support and facilitate the pattern matching process, allowing it to be developed as needed to achieve the best results.

b) Breakpoint Pattern Recognition: The closest or adjacent star methodology is used in this research's breakpoint pattern identification method, which is based on the concept of a star pattern recognition algorithm using a satellite star sensor [29]. The method's star pattern is replaced with a breakpoint pattern. The main catalog and the sub-catalogue are the two catalogs. The primary catalog contains IDs for all breakpoints, as well as two attributes: position in Cartesian coordinates and type (x,y). The sub catalog contains all of the neighboring breakpoint point IDs, as well as the distance between the breakpoints. The main idea behind this method is to compare each distance between the three observed breakpoint points to a catalog of previously defined space mappings.

V. Result and Discussion

A. Noise Data Model Design of RpLidar A1

The Rplidar A1 noise data model is made up of three conditions: 0-60, 60-100, and 100-225 cm at the point nearest to the sensor and the wall. The first condition is approximated by 6-order polynomial regression, while the values of the variables in the equation are approximated by 3rd order polynomial interpolation. The regression used is appropriate to utilize 2nd order polynomials in the range of 60 to 225, but the range of 60 to 100 values of the element variables is produced by 3rd order interpolation. The range of 100 to 225 is then estimated using linear interpolation.

B. Adaptive Line Tracking Breakpoint Detector (ALTBD)

The measurement inaccuracy (σ_{offset}) percentage value utilized is 0.04. This value is the average of the percentage of errors obtained from data collection results in the first stage of the research. The tolerance value as the corner breakpoint limit (Dth_{corner}) is 23 cm, which is half the door distance of 46 cm. Fig. 9 depicts one of the ALTBD module's results and Fig. 10 depicts one of the ABD IEPF module's results.

The ALTBD results show the detection of several corner breakpoints generated by the 18 mm thick side wall. Because the detected side is close to the sensor, the number of scanning points created is sufficient to make a segment with more than four points. Breakpoints that should be designated corner breakpoints by the IEPF method are regarded terminal breakpoints, in contrast to ABD IEPF. This is due to the fact that the required threshold distance is not met on that side.

C. Adaptive Breakpoint Detector (ABD) and IEPF

The threshold value equation utilized is consistent with the results of the Weerakoon et al. [20] research. The residual error is 0.0038, whereas the angle (λ) is 8º. The extraction line used is an IEPF with a limit value of 46 cm. The intersection of the two lines generated by the IEPF two-segment linear regression is used to determine the corner breakpoint.

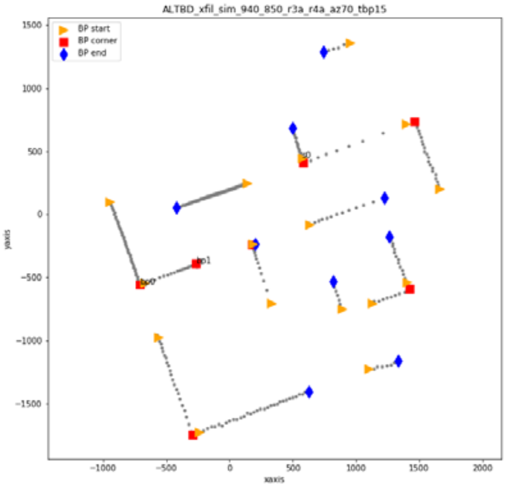


Fig. 9. Breakpoint Detector Results of ALTBD.

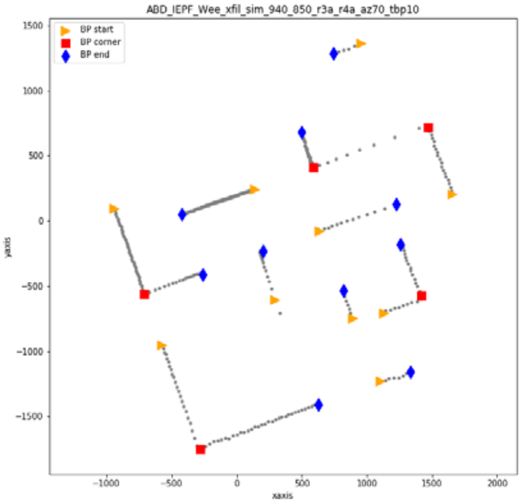


Fig. 10. Breakpoint Detector Results of ABD IEPF.

D. Testing

Fig. 11 depicts a comparison of the computing times of the three Breakpoint Detector methods. The comparison of the ALTBD, ABD IEPF, and LT methods reveals that the LT method computes faster than the ALTBD and ABD IEPF methods. The average computation time for the LT technique is 59,87 ± 1,94 ms, whereas the ALTBD and ABD IEPF methods require 101.61 ± 2.63 ms and 175.4 ± 10.13 ms, respectively. Although LT has the shortest calculation time, the resulting corner breakpoints are not as precise as ALTBD and ABD IEPF. If the accuracy of the LiDAR sensor's position in the room is a major priority in robot localization, this will be a disadvantage of the LT method. The ALTBD method has the second fastest computation time. Despite the fact that the ALTBD calculation is more sophisticated, each data point is only processed once. Unlike the IEPF ABD, the data is processed at least twice in this method. This method processes data twice, once using the ABD method and once using the IEPF method. The theory behind the significant difference in computational time is that the LT, ABD, and ALTBD methods are linear search algorithms with a time complexity of O(n), whereas the IEPF method is a linear recursive algorithm with a time complexity of $T(n) = 2T(n/2) + O(n)$, with the worst-case complexity being O(n^2). Table I presents the detailed results of the computational time comparison.

According to the test results, the IEPF's disadvantage is the recognition of an inaccurate corner breakpoint. This takes the form of a four-sided space, as illustrated in Fig. 10. This is due to the fact that the IEPF seeks the greatest deviation from the line connecting the starting and finishing points. Furthermore, there is a lot of noise at that point. This type of room is said to be unfavorable for detecting corner breakpoints using the IEPF method.

E. Breakpoint Extraction

The results of the tests show that the IEPF method has weaknesses, including the inaccuracy of recognizing corner breakpoints in the form of a four-sided room, as illustrated in Fig. 12. Table II indicates that the ALTBD method detects corner breakpoints with only two errors out of 43 position samples tested, but the ABD IEPF detects them with ten. When compared to ALTBD and ABD IEPF, the LT method has the most detection errors. If there are only two corner breakpoints on one side of the wall, the accuracy of the corner breakpoint is declared correct. It is still considered a corner breakpoint detection error if there are more than two corner breakpoints, even if they are close to one other.

The IEPF method is inaccurate in recognizing corner breakpoints because it looks for the greatest difference between the start and finish lines. Furthermore, there is a lot of noise at that point. When using IEPF, this type of room will result in an inaccurate corner breakpoint. In addition, ALTBD generates an inaccurate a corner breakpoint. Because the erroneous corner breakpoint is typically located distant from the LiDAR sensor, it has little effect on the location results of the Breakpoint Pattern Recognition, as illustrated in Fig. 13. However, with the spatial pattern depicted in Fig. 12, the resulting corner breakpoint position in the IEPF is not quite close to the LiDAR sensor. The Breakpoint Pattern Recognition method will not produce the right position if these two spots are not actual corner breakpoints. Some corner breakpoint detection problems in the LT method occur when the incorrect position is close to the sensor and the distance to the other reference corner breakpoint is within the Breakpoint Pattern Recognition tolerance limit, as illustrated in Fig. 14. As a result, the Breakpoint Pattern Recognition procedure is unable to locate a pattern match in the database.

F. Breakpoint Pattern Recognition

There are three stages to the breakpoint pattern recognition process. The first step is to look for a pattern match for the three initial reference points, which are the three breakpoints nearest to the sensor. The results of the search throughout the sub-catalogues are saved. Step two must be completed if there is more than one possibility. The second stage is to find a match against the original three reference points on the remaining corner breakpoints and terminal breakpoints in all sub-catalogues. The match of the nearby breakpoint points with the most points that determines the three initial patterns is eligible for selection. If stage two yields two or more results, stage three will be carried out. Stage three follows the same principles as stage two, with the exception of the matching point. This is the last of the breakpoints on the list (points that form a line). If step three is completed, all breakpoints have been verified.

TABLE I. COMPUTATIONAL TIME RESULTS

Location	**Num. sample**	**Computational time in ms**		
		ALTBD	***ABDiepf***	***LT***
Room 1	9	103,3 ± 8,0	198,9 ± 11,0	62,4 ± 0,0
Room 2	8	103,8 ± 8,0	208,6 ± 8,0	58,5 ± 7,2
Room 3	10	99,5 ± 8,0	157,9 ± 15,5	56,5 ± 8,0
Room 4	8	101,4 ± 8,0	169,0 ± 11,7	62,4 ± 0,0
Street	8	100,3 ± 8,0	135,9 ± 21,5	60,2 ± 5,9

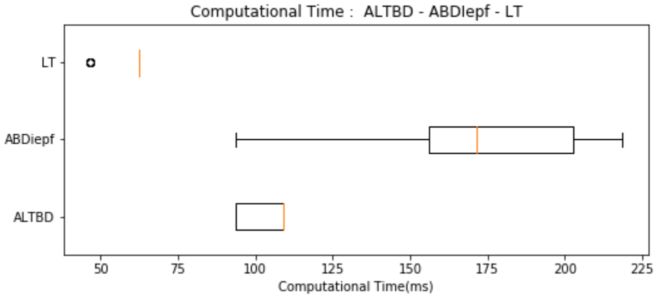


Fig. 11. Comparison of ALTBD, ABD IEPF and LT Computational Times in a Boxplot.

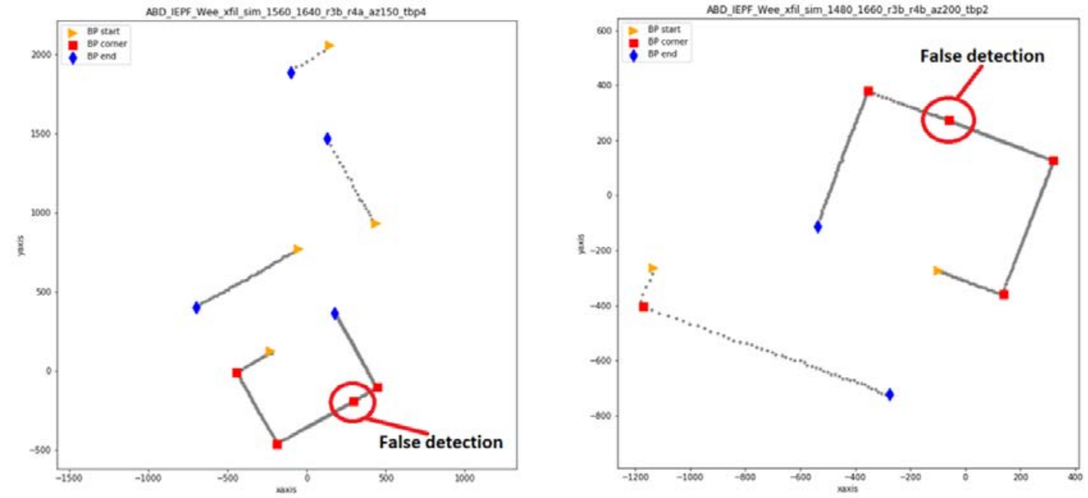


Fig. 12. IEPF Corner Breakpoint Detection.

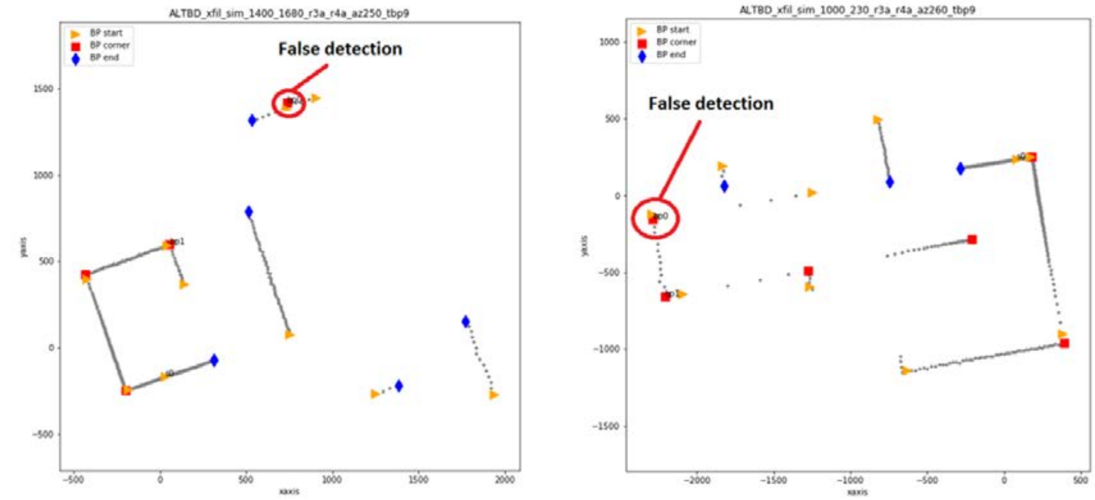


Fig. 13. ALTBD Corner Breakpoint Detection.

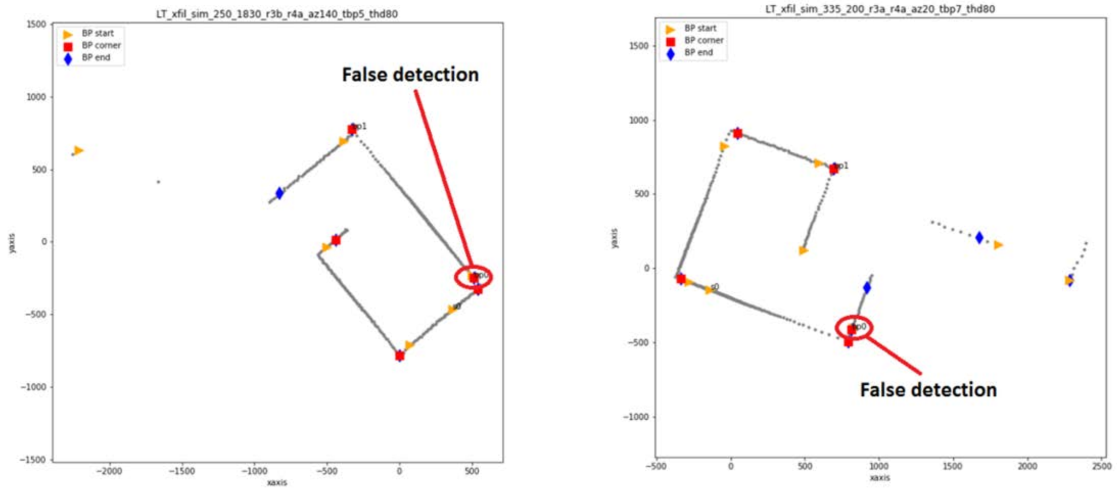


Fig. 14. LT Corner Breakpoint Detection.

TABLE II. CORNER BREAKPOINT DETECTION RESULTS

Location	**Num. sample**	**Corner breakpoint**					
		ALTBD		***ABDiepf***		***LT***	
		True	***False***	***True***	***False***	***True***	***False***
Room 1	9	33	0	34	3	36	4
Room 2	8	32	0	32	0	32	6
Room 3	10	30	0	32	3	36	3
Room 4	8	28	1	28	2	26	3
Street	8	34	1	32	2	45	7

Furthermore, the major catalog should be summarized in order to limit the number of candidate positions generated. In fact, one wall has four sides: two long sides and two wide sides, or what is generally referred to as the thick side. The wall has four breakpoints at first, which are then combined into two breakpoints with the condition that the wall thickness is included in the tolerance limit value. The main catalog, which originally had 40 points, was reduced to 19 points. Similarly, the sub-catalog averaged 19 nearby breakpoints at the end. The reduced number of neighbors in the reference point accelerates the matching process.

Fig. 15, 16, and 17 shows that the position generated by ALTBD is more accurate than ABD IEPF through the pattern recognition process using the closest breakpoint method. This is due to the fact that the reference points for position computations are only the two closest corner breakpoints. ALTBD produces the closest corner breakpoint more precisely than ABD IEPF.

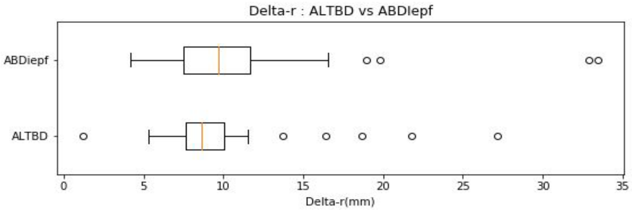


Fig. 15. Boxplot of the difference between ALTBD and ABD IEPF Position Errors.

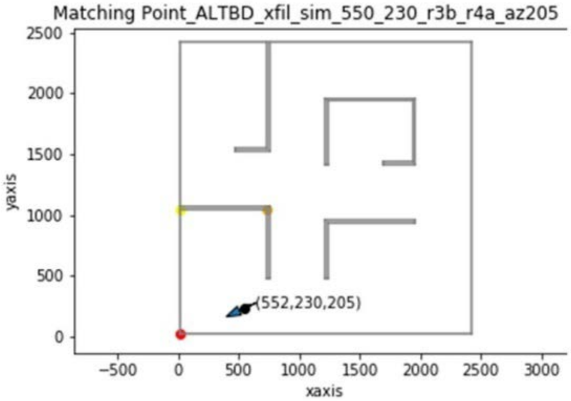


Fig. 16. Breakpoint Pattern Recognition Results of ALTBD.

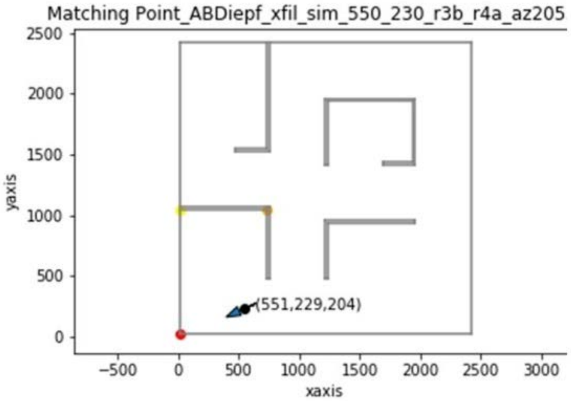


Fig. 17. Breakpoint Pattern Recognition Results of ABD IEPF.

VI. CONCLUSION

This research was successful in developing a new algorithm for breakpoint detection called the Adaptive Line Tracking Breakpoint Detector (ALTBD). This method modifies the Line Tracking (LT) and Adaptive Breakpoint Detector (ABD) algorithms by introducing a new threshold area in the shape of an ellipse, resolving corner breakpoint detection more adaptive and fasting. Algorithm testing was done by apply the ALTBD and ABD algorithms with Iterative End Point Fit (ABD IEPF) to detect the position of the robot in the room.

The results of the tests prove that the ALTBD computation time is faster in detecting corner breakpoints than the ABD IEPF method. The average computation time for the ALTBD method is 101.61 ± 2.63 ms, while the ABD IEPF is 175.4 ± 10.13 ms. The corner breakpoint detection error in the ALTBD method is only two errors out of 43 position samples, whereas the ABD IEPF method has ten detection errors. Furthermore, the ALTBD method is more accurate in determining the position of the robot than the ABD IEPF method, with a distance difference of 9.72 ± 1.55 mm, instead of 11.2 ± 2.14 mm in the ABD IEPF.

ACKNOWLEDGMENT

The authors would like to thank the Head of the Satellite Technology Research Center (BRIN) and Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) IPB University for their support in completing this research. This research is partially sponsored by the Ministry of Research and Technology of the Republic of Indonesia (RISTEK-BRIN) through Penelitian Dasar Unggulan Perguruan Tinggi (PDUPT) research grant number 3626/IT3.L1/PT.01.03/P/B/2022.

REFERENCES

[1] R. Gonzalez, F. Rodiguez, J. L. Guzma, and M. Berengguel, “Comparative study of localization techniques for mobile robots based on indirect kalman filter”, International Symposium on Robotics (ISR), pp:253–258. Switzerland. http://www.ual.es/~rgs927/papers/r amon-gonzalez-isr09.pdf, 2009.

[2] J. Borenstein, H. R. Everett, L. Feng, and D. Wehe, (1997) “Mobile robot positioning: sensors and techniques”, Journal of Robotic Systems, 14(4), pp. 231–249. doi: 10.1002/(SICI)1097-4563(199704)14:4<231::AID-ROB2>3.0.CO;2-R.1997.

[3] J. Park, M. Choi, Y. Zu, and J. Lee, “Indoor localization system in a multi-block workspace”, *Robotica*, 28(3), pp. 397–403. doi: 10.1017/S0263574709005712, 2010.

[4] N. L. Doh, H. Choset, and W. K. Chung, “Relative localization using path odometry information”, Autonomous Robots, 21(2), pp. 143-154. doi: 10.1007/s10514-006-6474-8, 2006.

[5] M. Faisal and H. ElGibreen, "Adaptive Self-Localization System for Low-Cost Autonomous Robot", 7th International Conference on Control, Automation and Robotics (ICCAR),Singapore, DOI:10.1109/ICCAR52225.2021.9463494, 2021.

[6] J. H. Kim, and P. H. Seong, “Experiments on orientation recovery and steering of autonomous mobile robot using encoded magnetic compass disc”, IEEE Transactions on Instrumentation and Measurement vol. 45, no. 1, pp. 271-273. https://doi.org/10.1109/19.481346, 1996.

[7] P. Artiemjew dan K. Ropiak, "Robot Localization in the Magnetic Unstable Environment", 2019 Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, doi: 10.1109/IRC.2019.00105, 2019.

[8] K. T. Song, and Y. H. Suen, “Design and implementation of a path tracking controller with the capacity of obstacle avoidance”, Journal of Control Systems and Technology , Vol. 4, No. 3, pp.151-160. Taipei, Taiwan, 1996.

[9] T. D. Kwon, and J. S. Lee, “A stochastic map building method for mobile robot using 2-d laser range finder”, 7th Autonomous Robots, 1-18. Kluwer Academic Publishers. Boston. https://doi.org/10.1023/ A:1008966218715, 1999.

[10] C. C. Hsu, H. C. Chen,C. C. Wong, and C. Y. Lai, “Omnidirectional Ultrasonic Localization for Mobile Robots”, Sensors and Materials, Vol. 34, No. 2, https://doi.org/10.18494/SAM3419, Tokyo, 2022.

[11] B. Tao, H. Wu, Z. Gong, Z. Yin, and H. Ding, “An RFID-Based Mobile Robot Localization Method Combining Phase Difference and Readability”, IEEE Transactions on Automation Science and Engineering, Vol. 18, doi: 10.1109/TASE.2020.3006724, 2021.

[12] P. Jensfelt, and H. Christensen, “Laser based position acquisition and tracking in an indoor environment”, Proc. of the Intl. Symposium on Robotics and Automation, 1(May), pp. 331–338, Available at: https://www.researchgate.net/publication/238648464, 1998.

[13] G. A. Borges, and M. J. Aldon, “Line extraction in 2D range images for mobile robotics”, Journal of Intelligent and Robotic Systems: Theory and Applications, 40(3), pp. 267–297. doi: 10.1023/B:JINT.0000038 945.55712.65, 2004.

[14] A. Siadat, A. Kaske, S. Klausmann, M. Dufaut, and R. Husson, “An optimized segmentation method for a 2D laser-scanner applied to mobile robot navigation”, IFAC Proceedings Volumes, 30(7), pp. 149–154. doi: 10.1016/s1474-6670(17)43255-1, 1997.

[15] K. J. Lee, “Reactive navigation for an outdoor autonomous vehicle”, Master Thesis. University of Sydney, Department of Mechanical and Mechatronic Engineering. Australia, 2001.

[16] K. C. J. Dietmayer, J. Sparbert, and D. Streller, ”Model based object classification and object tracking in traffic scenes from range images”, In: Proceedings of IV IEEE Intelligent Vehicles Symposium, Tokyo, 2001.

[17] S. Santos, J. E. Faria, F. Soares, R. Araujo, and U. Nunes, “Tracking of multi-obstacles with laser range data for autonomous vehicles”, In: Proc. 3rd National Festival of Robotics Scientific Meeting (ROBOTICA), pp. 59-65, 2003.

[18] N. Certad, R. Acuna, A. Terrones, D. Ralev, J. Cappelletto, and J. C. Grieco, ”Study and improvements in landmarks extraction in 2D range images based on an Adaptive Curvature Estimation”, Andean Region International Conference (ANDESCON) VI. Cuenca, Ecuador. https://doi.org/10.1109/Andescon.2012.31, 2012.

[19] S. Y.An, J. G.Kang, L. K. Lee, and S. Y. Oh, “Line segment-based indoor mapping with Salient Line Feature Extraction”, *Advanced Robotics*, 26(5–6), pp. 437–460. doi: 10.1163/156855311X617452, 2012.

[20] T. Weerakoon, K. Ishii, and A. A. F. Nassiraei, “Geometric feature extraction from 2D laser range data for mobile robot navigation”, *2015* IEEE 10th International Conference on Industrial and Information Systems, ICIIS 2015 - Conference Proceedings, pp. 326–331. doi: 10.1109/ICIINFS.2015.7399032, 2016.

[21] J. Dingyao, C. Jin, and X. Yuan, “An Extracting Method of Corner Points from Laser Sensor Readings”, Proceedings of the 37th Chinese Control Conference, Wuhan, China, doi:10.23919/ChiCC.2018.84825 34, 2018.

[22] V. Nguyen, A. Martinelli, N. Tomatis, and R. Siegwart, “A comparison of line extraction algorithms using 2D laser rangefinder for indoor mobile robotics”, in 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 1929–1934. doi: 10.1109/IROS.2005.1545234, 2005.

[23] J. Vandorpe, H. V. Brussel, and H. Xu, “Exact dynamic map building for a mobile robot using geometrical primitives produced by a 2D range finder”, Proceedings of IEEE International Conference on Robotics and Automation. pp.901-908. Minnesota, USA. https://doi.org/10.1109/RO BOT.1996.503887, 1996.

[24] G. A. Borges, and M. J. Aldon, “A Split-and-Merge Segmentation Algorithm for Line Extraction in 2-D Range Images”, Proceedings 15th International Conference on Pattern Recognition, ICPR-2000. Barcelona, Spain. https://doi.org/10.1023/B:JINT.0000038945.55712.6 5, 2000.

[25] R. O. Duda RO and P. E. Hart, “Pattern Classification and Scene Analysis”, John Wiley and Sons, Hoboken, 1973.

[26] M. Fischler and R. Bolles, “Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography”, Communications of the ACM, 24(6):381–395. https://doi.org/10.1145/358669.358692, 1981.

[27] D. A. Forsyth and J. Ponce, “Computer Vision: A Modern Approach”, Prentice Hall, 2003.

[28] KRPAI 2019 Guide, [online] Available: https://kontesrobotindonesia.id /data/2019/Panduan_KRPAI2019.pdf.

[29] M. A. Saifudin, and R. H. Triharjanto, ”*Algoritma pengenalan pola bintang untuk deteksi posisi bintang pada star sensor satelit LAPAN*”, *Jurnal Teknologi Dirgantara*, 8(1), pp. 36–42, Indonesia, 2010.

Alignment of Software System Level with Business Process Level: Resolving Syntactic and Semantic Conflicts

Samia Benabdellah Chaouni[1]
Department of Mathematics and Computer Science
Faculty of Sciences Ain Chock, Hassan II University
Casablanca, Morocco

Maryam Habba[2], Mounia Fredj[3]
AlQualsadi Research Team
ENSIAS, Mohammed V University in Rabat
Rabat, Morocco

***Abstract*—Information systems help organizations manage their entities with innovative technologies. These entities are often very different in nature. In this paper, we consider a business process level based on a set of Business Process Model and Notation (BPMN) models and a software system level based on a Unified Modeling Language (UML) class diagram. The differences between these entities make them difficult to align. In addition, an organization's BPMN models may be designed by different teams, which can cause syntactic and semantic heterogeneities. We present the first step of our proposed approach for aligning a software system level with a business process level without conflict (redundancy and lost information). Syntactic and semantic rules based on ontologies and other resources for comparing BPMN models are described, as well as a process for transforming BPMN models into UML model.**

***Keywords*—Information system alignment; business process; software system; Business Process Model and Notation (BPMN); Unified Modelling Language (UML); class diagram; ontology; semantic aspects**

I. Introduction

As organizations increase in number, competition between them intensifies. In order to compete, organizations adopt innovative strategies, seek high quality human resources, follow best practices, and use the most efficient technological tools. Developing an efficient and cost-effective information system is crucial for an organization's ability to compete. For this reason, alignment is vital for organizations. Indeed, alignment can provide solutions to problems associated with the diverse changes that may occur in an organization's entities. Business/IT alignment has been the topic of several previous studies [1]–[5]. The approach proposed in this paper is relevant in various situations. An information system with aligned levels may experience changes in one of its levels due to a revision of goals or other factors. This causes the levels to become misaligned. In addition, different levels of an organization's information system may be modelled by different teams. Each team perceives the system from a different perspective, which can also result in misaligned levels. Similarly, when two organizations with levels of different natures merge, the levels of the resulting information system will likely be misaligned as well. In this case, the resulting information system will contain misaligned levels. Further, most organizations include several BPMN models at the business process level, all built independently of each other. This can cause conflicts during alignment, due to heterogeneities between the models. In fact, existing approaches considering a set of BPMN models align models syntactically. However, they only test identity. They do not detect other correspondences, such as inclusions, abbreviations and acronyms. Moreover, none of these studies takes semantic aspects into account (synonymy, homonymy or hyponymy). These heterogeneities can cause conflicts when models are aligned and may introduce problems and inconsistencies in the resulting UML model. Indeed, if alignment approaches based on identity only consider two BPMN elements to be different when they are equivalent, they will introduce a false difference and, therefore, redundancy in the UML model. Worse, considering two elements as equivalent when they are different will introduce a false equivalence, resulting in information loss in the resulting UML model. A generalisation relationship between hyponymic elements is also missing in the UML model.

In all these situations, an effective alignment approach is required to obtain a successful information system. Additionally, it is necessary to ensure that alignment is achieved without conflicts (i.e. without loss or redundant information).

The different approaches described in previous studies concerning business process and software system levels are analysed in this paper. Following that analysis, we propose an alignment approach that resolves syntactic and semantic conflicts. The first step of the proposed approach includes a set of rules for comparing BPMN models to detect equivalencies and differences as well as a transformation process for converting a set of BPMN models into a UML class diagram [6]. The second step provides a method for preserving software system level information.

This paper is organised as follows: Section II presents the background of the topic; it introduces the concept of alignment and ontology. Section III provides a brief overview of related work, while the proposed approach is presented in Section IV. Section V presents a case study. Finally, the conclusion outlines our objectives for future work.

II. BACKGROUND

A. The Concept of Alignment

In the literature, various expressions have been used to describe the term alignment. Chan [7] refers to alignment by fit and synergy. Henderson and Venkatraman [8] use the terms fit, integration, and interrelationships. Reich and Benbasat [9] use the word linkage. For Ciborra [10], alignment is defined as a bridge. Smaczny [11] describes it as fusion. Luftman [12] employs the term harmony, while Nickels [13] uses congruence. Alignment between business and IT is defined by Ullah and Lai [5], as “the optimized synchronization between dynamic business objectives/processes and respective technological services provided by IT”. According to Luftman [12], business-IT alignment concerns the application of IT in a timely and a suitable manner, in harmony with business strategies, goals and needs. For Luftman [12], this definition considers the way IT is aligned with the business, and the way the business should or could be aligned with IT. In present work, we define alignment of a target level with a source level as a method that ensures the continued operation of the target level, while remaining suitable to the source level.

B. Ontology

Ontology is defined as an “explicit specification of a conceptualization” [14]. Domain ontologies are ontologies built on a particular knowledge domain. There are many domain ontologies such as MENELAS (in the medical domain) [15] and TOVE (in the business management domain) [16]. The domain ontology is a semantically rich model (it can express equivalence, inverse, disjunction, symmetry, transitivity, etc.), and is defined as an exhaustive list of concepts (ontology class) and relations between these concepts describing a particular domain (e.g. Medicine, Business, E-Government).

III. RELATED WORK

In a previous work [17], we proposed a pattern system as a guideline, to help organizations apply the alignment. In addition, a systematic literature review was conducted [18] to present various approaches to the alignment of business requirement, business process and software system levels, that use different modelling languages.

In this study, we focus on UML and BPMN languages, because they are standards defined by the Object Management Group (OMG). More precisely, we focus on a business process level modelled by BPMN and a software system level modelled by a UML class diagram.

BPMN and class diagrams are subjects of interest in different approaches. Amr et al. [19] propose an MDA approach for transforming a BPMN source model into a UML class diagram, using a set of transformation rules. Brdjanin et al. [20] present an approach for the automated generation of a conceptual database model represented by a UML class diagram from a single BPMN model. Brdjanin et al. [21] take a set of business process models into account. Khlif et al. [22] describe an approach to transform a business process model into a class diagram, based on aspects descriptions. Rhazali et al. [23] suggest a set of rules for transforming a BPMN model into a use case, state and class diagrams. Cruz et al. [24] propose an approach to obtain a data model from a business process model. Cruz et al. [25] present rules to transform a set of business process models into a data model. Kriouile et al. [26] describe an approach to transform a BPMN model into a domain class model. Bousetta et al. [27] propose an approach for building a domain class diagram based on a BPMN model, using a set of business rules.

In organizations, models of both levels usually exist. An analysis of existing approaches makes it clear that all existing approaches propose transformation from the source level into the target level. However, an approach-based transformation is not always sufficient to apply alignment when business process and software system models exist. In fact, these approaches can cause a loss of information. Fig. 1 presents the result of applying one of the existing approaches. M1 represents the business process level model, while M2 represents the software system level model.

Let X, Y and Z be three models belonging to business process level (M1). X contains elements A, F, K and I. Y contains elements B and J, while Z contains elements G and H. Model M2 contains elements D and E. The existing approaches can generate a new UML class diagram (M2’), containing T(A), T(F), T(K), T(I), T(B), T(G) and T(H), which represent the results of the transformation of elements A, F, K, I, B, G and H respectively.

We note that M2’ differs from model M2. Therefore, information D and E associated with the existing UML class diagram, will be lost after alignment (TABLE I, column 2). None of the existing approaches consider all BPMN elements frequently used in organizations (complete metamodel), which provide a detailed description of the models and which belong to the latest BPMN 2.0.2 specification [28], such as all task types or all data types (TABLE I, column 3). The majority of approaches take a single model at the source level into consideration. Only two existing approaches ([21] and [25]) have achieved transformation using a set of BPMN models as a source (TABLE I, column 4).

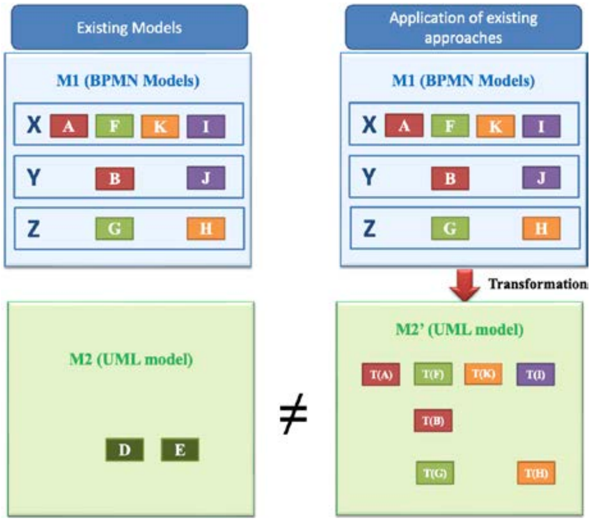


Fig. 1. Application of Existing Transformation Approaches.

TABLE I. SYNTHESIS OF APPROACHES

Ref.	Preserving information	Complete BPMN MM	Set of BPMN models	Syntactic aspect				Semantic aspect		
				Id	Inc	Ab	Ac	S	Hp	Hm
[19]	-	-	-	-	-	-	-	-	-	-
[20]	-	-	-	-	-	-	-	-	-	-
[21]	-	-	x	x	-	-	-	-	-	-
[22]	-	-	-	-	-	-	-	-	-	-
[23]	-	-	-	-	-	-	-	-	-	-
[24]	-	-	-	-	-	-	-	-	-	-
[25]	-	-	x	x	-	-	-	-	-	-
[26]	-	-	-	-	-	-	-	-	-	-
[27]	-	-	-	-	-	-	-	-	-	-

Because models X, Y and Z are usually created independently, they include many types of heterogeneity. All approaches considering a set of BPMN models align models elements syntactically. However, they only test identity: the string equality of elements. They do not detect other correspondences, such as inclusions (e.g. element A, “Medical Office” in X and element B, “Office” in Y), abbreviations (e.g. “Qty” and “Quantity”) and acronyms (e.g. "UOM" and "Unit Of Measure") (TABLE I, column 5). Moreover, none of these studies take semantic aspects into account (TABLE I, column 6). They do not detect synonymy (e.g. element F, "Doctor" in X and element G, "Medical Practitioner" in Z), homonymy (e.g. element I, "Invoice" (of Patient) in X, and element J "Invoice" (of Supplier) in Y), or hyponymy (a semantic relationship between terms where the meaning of one is included in another, more general term) (e.g. element K, “Patient” in X and element H, “Diabetic” in Z).

As shown in Fig. 1, these heterogeneities can cause conflicts when models are aligned and may introduce problems and inconsistencies in the resulting UML model. Indeed, if alignment approaches based on identity only consider two BPMN elements (belonging to X and Y) to be different when they are equivalent, they will introduce a false difference and therefore redundancy in the UML model. For example, we can find both T(A) “Medical Office” and T(B) “Office” in the UML model. Worse, considering two elements (belonging to X and Y) as equivalent when they are different will introduce a false equivalence, resulting in information loss in the resulting UML model. For example, only T(I) “Invoice” (of Patient) can be found in the UML model, and not “Invoice” (of Supplier). A generalisation relationship between hyponymic elements is also missing in the UML model. For example, the generalisation relationship between element T(K) “Patient” and element T(H) “Diabetic”.

We synthesize the existing approaches in TABLE I according to the following criteria: preserving information, considering complete BPMN metamodel (MM), considering a set of BPMN models, considering syntactic aspects (we note syntactic comparison, the comparison strings of model elements’ letters. It indicates if the approach detects identity (Id), inclusions (Inc), abbreviations (Ab) and acronyms (Ac)), and considering semantic aspects (we note semantic comparison, the comparison of the meaning associated with the model’s elements). It indicates if the approach detects synonymy (S), hyponymy (Hp) and homonymy (Hm). In TABLE I, “**x”** indicates that a criterion is considered.

This analysis of existing approaches reveals the need for an alignment approach that preserves existing information, uses a set of BPMN models and considers most used BPMN elements. Moreover, our objective is to propose how to find real equivalences and real differences and how to get UML result model without conflicts.

IV. PROPOSED APPROACH

A. *Overview of the Proposed Approach*

The aim of the proposed approach is to align the software system level with the business process level, without losing information and without conflict.

We present an alignment system in Fig. 2 that takes a set of BPMN models as input and outputs the resulting UML model. It encompasses two steps:

1) *Step 1: Comparison and transformation*

a) Comparison of BPMN models: Our goal is to provide a semantic comparison approach that also integrates syntactic aspects. The model comparison subsystem takes BPMN models as input and returns (1) a comparison table containing the correspondence between elements (equivalent, different and hyponymic elements) and (2) isolated elements (those without equivalent, homonym or hyponym in another model). It is a syntactic and semantic rules-based system (presented in section IV.C.1), driven by a comparison process. We use strategies based on semantic properties to take semantic aspects into account. Therefore, our system refers to a domain ontology that will provide semantically relevant information for decision-making during the comparison (example: two ontology classes are semantically equivalent, two elements are hyponyms presented by an ontology class and its ontology subclass, etc.).

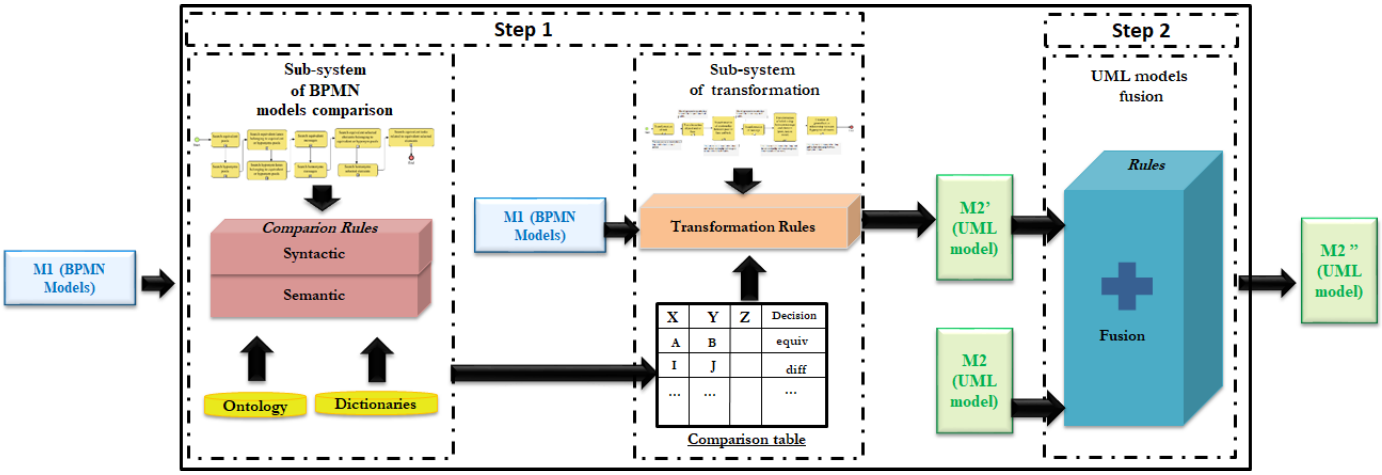


Fig. 2. Representation of Our Alignment Approach.

In addition, our system relies on several other resources to complete these comparisons. We use an acronym dictionary[1], an abbreviation dictionary[2], and a dictionary of synonyms[3].

b) Transformation of BPMN models into a UML model: This subsystem consists of the application of rules to transform a set of BPMN models (M1) into a generated UML class diagram (M2'), based on a comparison table and following a transformation process. Transformation rules based on MDA are presented in our previous work [29]. In the present work, we update the transformation process from BPMN models into a UML model, considering syntactic and semantic aspects (presented in section IV.C.2)).

By applying the rules of syntactic and semantic comparison, we can detect real equivalences and real differences, and thus obtain a UML result model without conflicts (Fig. 3). By identifying F and G as semantically equivalent, we are left with only one element, T(F), and therefore do not have redundancy in UML diagram M2'. Additionally, identifying A and B as syntactically equivalent results in only one element T(A), and thus no redundancy in M2'. Identifying I and J as homonyms produces two result elements, T(I) and T(J), in M2'. Finally, identifying K and H as hyponyms produces a generalisation relationship between the two resulting elements T(K) and T(H) in M2'.

2) Step 2: Fusion: This step consists of creating a fusion between the UML class diagram (M2') generated in step 1 and the existing UML class diagram (M2). We have previously demonstrated the results of this phase; the comparison and fusion meta-models were published in [30]. This solution also takes into account the syntactic and semantic aspects of the two class diagrams elements.

The result is a final UML class diagram (M2"). By applying the two steps illustrated in Fig. 3, the target level is completed, as it contains the information related to the existing class diagram (M2) as well as the information related to the generated class diagram (M2').

B. Example of BPMN Models

This approach can be applied to several BPMN models such us collaboration or process diagrams. In this section, we present a set of BPMN collaboration diagrams. To illustrate our approach, we use an example representing the business process level of a medical field.

A collaboration diagram can contain several elements: pool, lane, event, task, gateway, message flow, message, sequence flow, data, data association, artifact and association. A pool can refer to a process or can be a black box. A lane is a sub-partition within a pool; it can contain a set of events (facts that occur during the process), tasks (atomic work performed in a process), or gateways (controlling the convergence or divergence of flows in a process). A task can appear as a send task, a receive task, a service task, a user task, a manual task, a business rule task, or a script task. A message flow may contain a message (in the present work, we suppose that a message flow contains a message), and links source and target elements (pools, events, or tasks). A sequence flow connects a source and a target element (events, tasks, or gateways). Data provides information about what tasks need to be performed and/or what they produce. There are four types of data: data object, data store, data input or data output. A data association links source data or target data to a task. An artifact can be in the form of a group or a text annotation. It aims to provide more clarity to the process. An association links an artifact with a BPMN element.

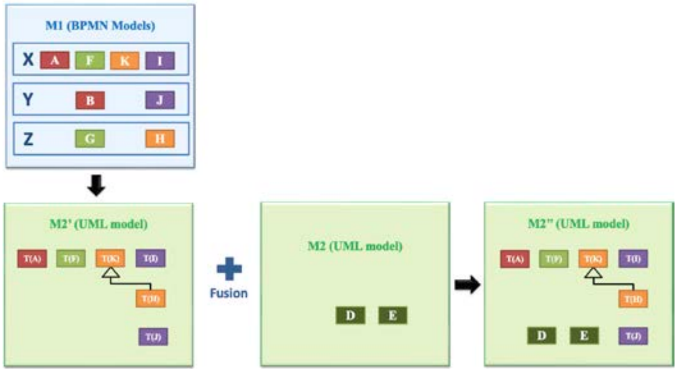


Fig. 3. Result Model without Conflicts or Information Loss.

[1] https://www.dictionary.com/e/acronyms/
[2] http://theleme.enc.sorbonne.fr/dico.php
[3] http://wordnet.princeton.edu/

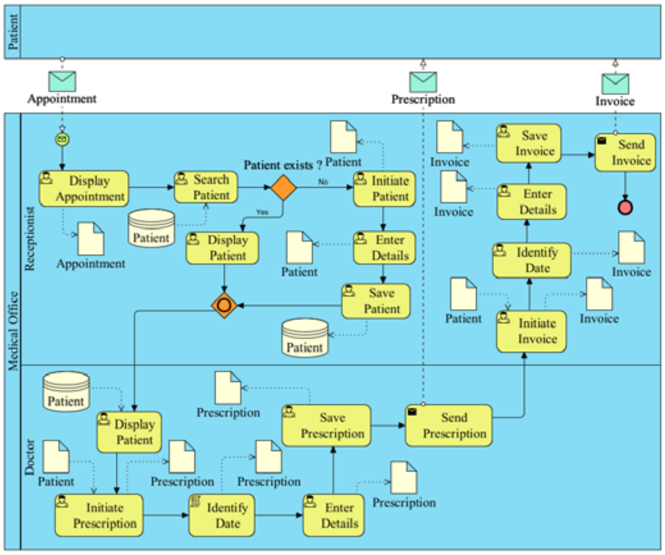


Fig. 4. Model X - Patient Consultation in a Medical Office.

Fig. 4 can illustrate the collaboration diagram (model X) for a patient consultation in a medical office. It is composed of two pools: “Patient”, which is a black box and “Medical Office”, which contains the lanes “Receptionist” and “Doctor”. The diagram begins when a patient arrives at the medical office. The receptionist performs the first task, “Display Appointment”, which has a data object “Appointment” as output. The receptionist then searches for the patient in a patient list. If a patient file exists, the receptionist displays the patient’s information. If not, the receptionist initiates the patient file, adds details and saves the file. The doctor then displays the patient file and initiates a prescription. The date of the prescription is identified. Then, the doctor enters details, saves the file, and sends the prescription. Next, the receptionist initiates an invoice, and its date is identified. The receptionist then enters details, saves the file, and sends the invoice.

Fig. 5 can represent the second collaboration diagram (model Y) placing orders with suppliers. It contains two pools: “Supplier”, represented as a black box, and “Office” which contains two lanes (“Assistant” and “Doctor”). The first task is “Initiate Purchase Order” performed by the assistant. It has as an output the data object “Purchase Order”. Then the date of the purchase order is identified. Next, the assistant adds details, and saves the purchase order. Then, the doctor displays, modifies and validates the purchase order. Next, the assistant sends the purchase order and then receives the invoice from the supplier.

Fig. 6 can illustrate the third collaboration diagram (model Z), for monitoring patients with diabetes. It contains two pools: “Diabetic”, represented as a black box, and “Medical Office”, which contains two lanes (“Assistant” and “Medical Practitioner”). The first task is “Search Patient File” executed by the assistant. Its input is the data store “Patient File”. If the patient is diabetic, the patient file is exposed. If the patient needs an appointment, the assistant initiates the appointment, adds details, saves and sends the appointment to the diabetic. Then the medical practitioner exposes the appointment.

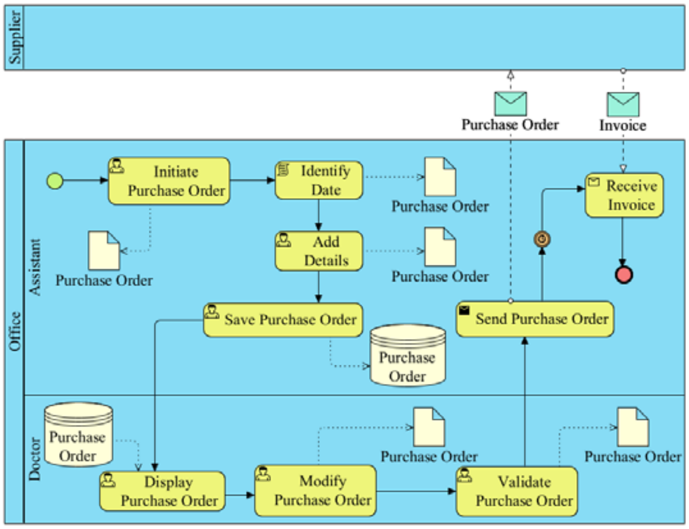


Fig. 5. Model Y - Ordering from Suppliers.

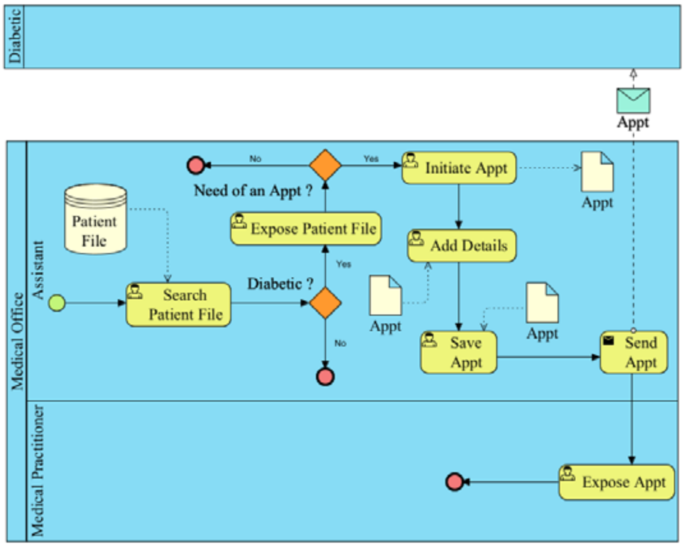


Fig. 6. Model Z - Monitoring Diabetics.

C. Comparison and Transformation Subsystems

In this section, we present the details of our approach. We first present the comparison of BPMN models, then the transformation of those BPMN models into a UML class diagram.

1) Comparison of BPMN models: Our goal is to compare BPMN models syntactically and semantically. To do so, we follow the model comparison process detailed in Section IV.C.1)a), which applies the BPMN element comparison rules described in Section IV.C.1)b).

a) Comparison process: We apply a comparison between each two models. To create a UML class diagram, we create classes before their operations. To do this, we first compare BPMN elements that will be transformed into classes: 1) *Pools, 2)* lanes, 3) messages and 4) data of non manual task and direct object (DO) of all types of task (except manual task) without data, send task that have a data in its input and receive task that have a data in its output). We call the fourth category of elements: “selected elements”. Next, we compare BPMN elements that will be transformed to operations: tasks related to equivalent selected elements.

The BPMN elements that will be transformed to associations, aggregations, attributes, and multiplicities [presented in Section IV.C.2)] are not concerned by comparison, because they are not named elements and therefore cannot be compared syntactically and semantically. The elements events, gateways, manual tasks, data related to manual tasks, sequence flows, and artifacts are not considered also in this comparison, because they will not appear in the resulting UML model (they do not have an equivalent in the UML class diagram).

In order to compare BPMN models, we follow the steps of the comparison process (Fig. 7), in this order: search equivalent pools, search hyponym pools, search equivalent lanes belonging to equivalent or hyponym pools, search hyponym lanes belonging to equivalent or hyponym pools, search equivalent messages, search homonym messages, search equivalent selected elements belonging to equivalent or hyponym pools, search homonym selected elements, and finally search equivalent tasks related to equivalent selected elements.

Pools are not concerned with homonymy because BPMN models are designed for the same domain. For example, if “office” is found in the first model meaning medical office, and "office" is also found in the second model, it will refer to a medical office there as well, and not, for example, a lawyer's office. The same is true for lanes: two homonymic lanes cannot belong to equivalent pools.

b) Comparison rules: We define a mathematical framework to express formally our approach. We present the comparison rules by the predicate language. So, we needed to express, in predicate logic, BPMN model (representing input of system), ontology and other resources (representing the system references). For that, we realised transformations, in the Model Driven Architecture (MDA) context:

i) The first transformation concerns BPMN model into logical model that generates a set of predicates representing BPMN elements to compare:

- Element(e,M): The element Pool or Lane “e” belonging to the model “M”.
- Pool(P,M): The pool “P” belonging to the model “M”.
- Lane (L,P,M) : The Lane “L” belonging to the pool “P”and the model “M”.
- Message (m, SourceP, TargetP,M): The message “m” relating the source Pool “SourceP” and the target Pool “TargetP”, belonging to the model “M”.
- SelectedElement(E, P, M) : The selected element “E”, belonging to the pool “P” and the model “M”.
- Task (T, in, out, L, P, M) : The task “T”, with the input data “in”, and the output data “out”, belonging to the Lane “L”, the pool “P” and the model “M”.

ii) The second transformation concerns OWL ontology (conform to an extract of OWL metamodel [31]) into logical model. The transformation generates a set of predicates representing OWL ontology, such as:

- equivalentOntoClass(C1, C2): The ontology classes “C1” and “C2” are equivalent.
- OntoSuperClassOf(C1, C2): The ontology class “C1” is subclass of the ontology class “C2”.

Other system references are presented as follows:

- DicAcronyms(elt1,elt2): “elt1” and “elt2” are acronyms.
- DicAbbreviation(elt1,elt2) :“elt1” and “elt2” are on abbreviation relation.
- DicSynonymy(elt1,elt2) :“elt1” and “elt2” are synonyms.

To complete rules expression, we define a set of facts from programming languages, such as equality of two strings, and inclusion of two strings:

- String(elt) : “elt” is a character string.
- InclusionString(s1,s2) : means that the character string “s1” is included in the character string “s2”.
- EqualString(s1,s2): The two strings “s1” and “s2” are equal.

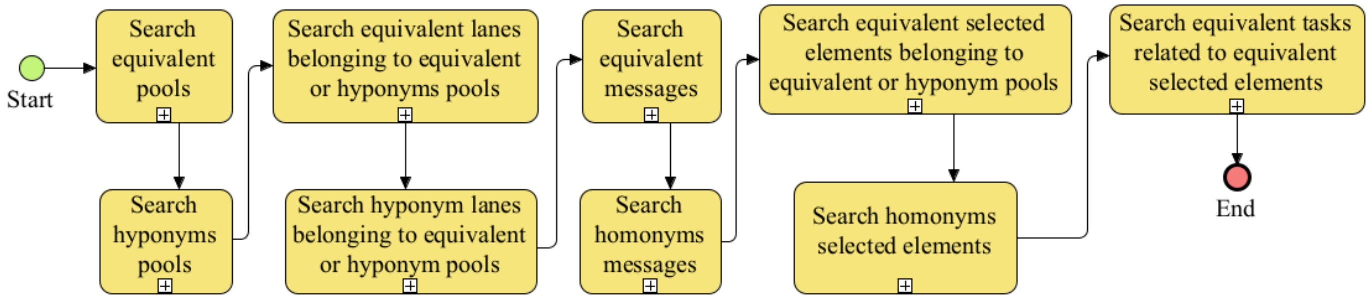

Start
Search equivalent pools
Search hyponyms pools
Search equivalent lanes belonging to equivalent or hyponyms pools
Search hyponym lanes belonging to equivalent or hyponym pools
Search equivalent messages
Search homonyms messages
Search equivalent selected elements belonging to equivalent or hyponym pools
Search homonyms selected elements
Search equivalent tasks related to equivalent selected elements
End

Fig. 7. Comparison Process.

Our system is based on a set of comparison rules which are characterized by a name and it is composed of a set of parameters (e.g. elements to compare, first model and second model) that have a name. A rule can call one or more other rules.

We present the comparison rules of models below.

- CR1 : Rule of syntactic identity of elements

$$\mathrm{Syntactic\_Identity}(\mathrm{elt}_i, \mathrm{elt}_j) \Leftrightarrow \mathrm{EqualString}\ (\mathrm{elt}_i, \mathrm{elt}_j)$$

Two elements elt_i and elt_j are syntactically identical if and only if they have the same name.

Example: The pool “Medical office” in X and the pool “Medical office” in Z.

- CR2: Rule of syntactic equivalence of elements

$$\begin{array}{l}
\mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_i, \mathrm{elt}_j) \Leftrightarrow [\mathrm{Syntactic\_Identity}\ (\mathrm{elt}_i, \mathrm{elt}_j) \lor \\
\mathrm{InclusionString}\ (\mathrm{elt}_i, \mathrm{elt}_j) \lor \mathrm{InclusionString}\ (\mathrm{elt}_j, \mathrm{elt}_i) \lor \\
\mathrm{DicAcronyms}(\mathrm{elt}_i, \mathrm{elt}_j) \lor \mathrm{DicAbbreviation}(\mathrm{elt}_i, \mathrm{elt}_j)]
\end{array}$$

Two elements elt_i and elt_j are syntactically equivalent if and only if they are identical, or if there is a relationship of inclusion, acronym or abbreviation of them (according to dictionaries).

Example: The pools “Medical Office” in X and “Office” in Y.

- CR3: Rule of semantic equivalence of elements

$$\mathrm{Equivalence\_Semantic\_Elements}(\mathrm{elt}_i, \mathrm{elt}_j) \Leftrightarrow$$

$$\begin{array}{l}
[(\mathrm{DicSynonymy}(\mathrm{elt}_i, \mathrm{elt}_j) \lor (\exists e\ \mathrm{String}(e) \land \\
\mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_i, e) \land \mathrm{DicSynonymy}(e, \mathrm{elt}_j)) \lor \\
(\exists c\ \mathrm{String}(c) \land \mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_j, c) \land \\
\mathrm{DicSynonymy}(c, \mathrm{elt}_i)) \lor (\exists e\ \mathrm{String}(e) \land \exists c\ \mathrm{String}(c) \land \\
\mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_j, e) \land \mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_i, c) \land \\
\mathrm{DicSynonymy}(e, c)))
\end{array}$$

$$\begin{array}{l}
\lor (\mathrm{equivalentOntoClass}(\mathrm{elt}_i, \mathrm{elt}_j) \lor (\exists e\ \mathrm{String}(e) \land \\
\mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_i, e) \land \mathrm{equivalentOntoClass}(e, \mathrm{elt})) \lor \\
(\exists c\ \mathrm{String}(c) \land \mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_j, c) \land \\
\mathrm{equivalentOntoClass}(c, \mathrm{elt}_i)) \lor (\exists e\ \mathrm{String}(e) \land \exists c\ \mathrm{String}(c) \land \\
\mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_j, e) \land \mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_i, c) \land \\
\mathrm{equivalentOntoClass}(e, c)))]
\end{array}$$

Two elements elt_i and elt_j are semantically equivalent if and only if one of these conditions is satisfied:

- elt_i (or a character string syntactically equivalent to elt_i) is synonymous with elt_j (or a character string syntactically equivalent to e_j), according to a synonym dictionary.
- there are two classes in the domain ontology with the same names of elt_i and elt_j (or a character string syntactically equivalent to elt_i and elt_j), which are equivalent.

Example: The lanes “Receptionist” in X and “Assistant” in Y.

- CR4: Rule of hyponyms of elements (Pool or Lane)

$$\begin{array}{l}
\mathrm{Hyponym\_Elements}(\mathrm{elt}_i, \mathrm{elt}_j, M_i, M_j) \Leftrightarrow [\mathrm{Element}(\mathrm{elt}_i, M_i) \land \\
\mathrm{Element}(\mathrm{elt}_j, M2) \land (\mathrm{OntoSuperClassOf}(\mathrm{elt}_i, \mathrm{elt}_j) \lor (\exists e\ \mathrm{String}(e) \land \\
\mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_i, e) \land \mathrm{OntoSuperClassOf}(e, \mathrm{elt}_j)) \lor \\
(\exists c\ \mathrm{String}(c) \land \mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_j, c) \land \\
\mathrm{OntoSuperClassOf}(\mathrm{elt}_i, c)) \lor (\exists e\ \mathrm{String}(e) \land \exists c\ \mathrm{String}(c) \land \\
\mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_i, e) \land \mathrm{Equivalence\_Syntactic\_Elts}(\mathrm{elt}_j, c) \land \\
\mathrm{OntoSuperClassOf}(e, c)))\ ]
\end{array}$$

An element (Pool or Lane) elt_i is a hyponym of an element elt_j if and only if, in a domain ontology, one ontology class with the same name of elt_i (or a character string syntactically equivalent to elt_i) is an ontology subclass of elt_j (or a character string syntactically equivalent to elt_j).

Example: The pool “Diabetic” in Z is a hyponym of the pool “Patient” in X.

- CR5: Rule of Equivalent Pools

$$\begin{array}{l}
\mathrm{Equivalence\_Pools}(P_i, P_j, M_i, M_j) \Leftrightarrow [\mathrm{Pool}(P_i, M_i) \land \mathrm{Pool}(P_j, M_j) \land \\
(\mathrm{Equivalence\_Syntactic\_Elts}(P_i, P_j) \lor \\
\mathrm{Equivalence\_Semantic\_Elements}(P_i, P_j)\ )]
\end{array}$$

Two pools P_i and P_j are equivalent if and only if they are syntactically or semantically equivalent.

Example: The pool “Medical Office” in X and the pool “Medical Office” in Z.

- CR6: Rule of Equivalent lanes

$$\begin{array}{l}
\mathrm{Equivalence\_Lanes}(L_i, L_j, M_i, M_j) \Leftrightarrow [\mathrm{Lane}(L_i, P_i, M_i) \land \mathrm{Lane}(L_j, P_j, M_j) \land \\
(\mathrm{Equivalence\_Syntactic\_Elts}(L_i, L_j) \lor \\
\mathrm{Equivalence\_Semantic\_Elements}(L_i, L_j))]
\end{array}$$

Two lanes L_i and L_j are equivalent if and only if they are syntactically or semantically equivalent.

Example: The lane “Doctor” belonging to the pool “Medical Office” in X and the lane “Medical Practitioner” belonging to the pool “Medical Office” in Z.

- CR7: Rule of Equivalent messages

$$\begin{array}{l}
\mathrm{Equivalence\_Messages}(m_i, m_j, M_i, M_j) \Leftrightarrow \\
[\mathrm{Message}(m_i\ \mathrm{SourceP}_i, \mathrm{TargetP}_i, M_i) \land \\
\mathrm{Message}(m_j, \mathrm{SourceP}_j, \mathrm{TargetP}_j, M_j) \land \\
(\mathrm{Equivalence\_Syntactic\_Elts}(m_i, m_j) \lor \\
\mathrm{Equivalence\_Semantic\_Elements}(m_i, m_j)) \land \\
((\mathrm{Equivalence\_Pools}(\mathrm{SourceP}_i, \mathrm{SourceP}_j, M_i, M_j) \lor \\
\mathrm{Equivalence\_Pools}(\mathrm{SourceP}_i, \mathrm{TargetP}_j, M_i, M_j) \lor \\
(\mathrm{Hymonym\_Elements}(\mathrm{SourceP}_i, \mathrm{SourceP}_j, M_i, M_j) \lor \\
\mathrm{Hymonym\_Elements}(\mathrm{SourceP}_j, \mathrm{SourceP}_i, M_i, M_j)) \lor \\
(\mathrm{Hymonym\_Elements}(\mathrm{SourceP}_i, \mathrm{TargetP}_j, M_i, M_j) \lor \\
\mathrm{Hymonym\_Elements}(\mathrm{TargetP}_j, \mathrm{SourceP}_i, M_i, M_j))) \land \\
(\mathrm{Equivalence\_Pools}(\mathrm{TargetP}_i, \mathrm{SourceP}_j, M_i, M_j) \lor \\
\mathrm{Equivalence\_Pools}(\mathrm{TargetP}_i, \mathrm{TargetP}_j, M_i, M_j) \lor \\
(\mathrm{Hymonym\_Elements}(\mathrm{TargetP}_i, \mathrm{SourceP}_j, M_i, M_j) \lor \\
\mathrm{Hymonym\_Elements}(\mathrm{SourceP}_j, \mathrm{TargetP}_i, M_i, M_j)) \lor \\
(\mathrm{Hymonym\_Elements}(\mathrm{TargetP}_i, \mathrm{TargetP}_j, M_i, M_j) \lor \\
\mathrm{Hymonym\_Elements}(\mathrm{TargetP}_j, \mathrm{TargetP}_i, M_i, M_j))))]
\end{array}$$

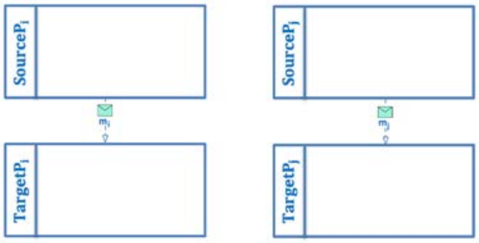


Fig. 8. Representation of Two Messages Mi and Mj.

Two messages m_i and mj (Fig. 8) are equivalent if and only if:

- they are syntactically or semantically equivalent.
- and their source and target pools are equivalent or hyponyms.

Example: The message “Appointment” connecting the pools “Medical Office” and “Patient” in X and the message “Appt” connecting the pools “Medical Office” and “Diabetic” in Z.

- CR8: Rule of homonym message

$$\begin{aligned}&\text{Homonym\_Messages}(m_i, m_j, M_i, M_j)\\ &\quad\Leftrightarrow [\text{Message}(m_i\ \text{SourceP}_i, \text{TargetP}_i, M_i)\\ &\quad\wedge \text{Message}(m_j, \text{SourceP}_j, \text{TargetP}_j, M_j)\\ &\quad\wedge \text{Equivalence\_Syntactic\_Elts}(m_i, m_j)\\ &\quad\wedge (\neg\text{Equivalence\_Pools}(\text{SourceP}_i, \text{SourceP}_j, M_i, M_j)\\ &\quad\vee \neg\text{Equivalence\_Pools}(\text{SourceP}_i, \text{TargetP}_j, M_i, M_j)\\ &\quad\vee \neg\text{Equivalence\_Pools}(\text{TargetP}_i, \text{SourceP}_j, M_i, M_j)\\ &\quad\vee \neg\text{Equivalence\_Pools}(\text{TargetP}_i, \text{TargetP}_j, M_i, M_j))]\end{aligned}$$

Two messages m_i and m_j (Fig. 8) are homonyms if and only if:

- they are syntactically equivalent.
- if at least one of their source or target pools are not equivalent.

Example: The message “Invoice” in X and the message “Invoice” in Y.

- CR9: Rule of equivalent selected elements.

$$\begin{aligned}&\text{Equivalence\_SelectetElements}(e_i, e_j, M_i, M_j)\ \Leftrightarrow\\ &[\text{SelectedElement}(e_i, P_i, M_i) \wedge \text{SelectedElement}(e_j, P_j, M_j) \wedge\\ &(\text{Equivalence\_Syntactic\_Elts}(e_i, e_j) \vee\\ &\text{Equivalence\_Semantic\_Elements}(e_i, e_j)) \wedge\\ &(\nexists m_i \text{Message}(m_i, \text{SourceP}_i, \text{TargetP}_i, M_i) \wedge\\ &\text{Equivalence\_Syntactic\_Elts}(e_i\ , m_i)\ \wedge\\ &\nexists m_j \text{Message}(m_j, \text{SourceP}_j, \text{TargetP}_j, M_j) \wedge\\ &\text{Equivalence\_Syntactic\_Elts}(e_j\ , m_j) \wedge\\ &\text{Homonym\_Messages}(m_i, m_j, M1, M2))]\end{aligned}$$

Two selected elements e_i and e_j are equivalent if and only if:

- they are syntactically or semantically equivalent.
- and there is no homonyms messages with the same name of selected elements.

Examples: The selected element “Patient” in X and the selected element “Patient File” are equivalent.

- CR10: rule of Homonyms selected elements

$$\begin{aligned}&\text{Homonym\_SelectetElements}(e_i, e_j, M_i, M_j)\ \Leftrightarrow\\ &[\text{SelectedElement}(e_i, P_i, M_i) \wedge \text{SelectedElement}(e_j, P_j, M_j) \wedge\\ &((\exists m_i \text{Message}(m_i, \text{SourceP}_i, \text{TargetP}_i, M_i) \wedge\\ &\text{Equivalence\_Syntactic\_Elts}(e_i\ , m_i)\ \wedge\\ &\exists m_i \text{Message}(m_j, \text{SourceP}_j, \text{TargetP}_j, M_j) \wedge\\ &\text{Equivalence\_Syntactic\_Elts}(e_j\ , m_j) \wedge\\ &\text{Homonym\_Messages}(m_i, m_j, M_i, M_j)) \vee\\ &((\text{Equivalence\_Syntactic\_Elts}(e_i, e_j) \wedge\\ &(\neg\text{Equivalence\_Pools}(P_i, P_j, M_i, M_j)) \vee\\ &\neg\text{Hymonym\_Elements}(P_i, P_j, M_i, M_j))]\end{aligned}$$

Two selected elements e_i and e_j are homonyms if and only if:

- there are homonyms messages with the same name of selected elements.
- or, they are syntactically equivalent and belong to non-equivalent and non-hyponym pools.

Example: The selected element “Invoice” in X and the selected element “Invoice” in Y.

- CR11: Rule of Equivalent tasks

$$\begin{aligned}&\text{Equivalence\_Tasks}(T_j, T_j, M_i, M_j) \Leftrightarrow [\text{Task}(T_i, \text{type}_{ti}, \text{dIN}_i, \text{dOUT}_i, L_i, P_i, M_i)\\ &\wedge \text{Task}(T_j, \text{type}_{tj}, \text{dIN}_j, \text{dOUT}_j, L_j, P_j, M_j)\\ &\qquad\wedge (\text{Equivalence\_Syntactic\_Elts}(T_i, T_j))\\ &\qquad\vee \text{Equivalence\_Semantic\_Elements}(T_i, T_j)]]\end{aligned}$$

Two tasks T_i and T_j are equivalent if and only if are syntactically or semantically equivalent.

Example: The task “Display Appointment” in X and the task “Display Appt” in Z.

c) Comparison table: We apply the process and the comparison rules on the examples of BPMN models presented in section IV.B, by referring to an ontology of the medical field as well as dictionaries. We choose OWL (Ontology Web language) ontology because it is a W3C[4] recommendation, and the metamodel OWL was defined by Ontology Definition Metamodel specification of OMG. To compare these models, we can find in ontology of medical domain several information. For a reason of space, we present bellow two information: Two equivalent ontology OWL classes, and the ontology OWL class “Patient” and its sub class “Diabetic”:

```
<owl:Class rdf:ID="Doctor">

<owl:equivalentClass rdf:resource="#Medical Practionnar"/>

   </owl:Class>

<owl:Class rdf:ID="Diabetic">

      <rdfs:subClassOf rdf:resource="#Patient"/>

</owl:Class>
```

[4] www.w3.org

TABLE II. COMPARISON TABLE

	X	Y	Z	Equiv syn/sem	Hypo	Hom	Decision	RCX	RCY	RCZ
Pools	Medical Office	Office	Medical Office	Yes	-	-	Equivalent	Medical Office	Medical Office	Medical Office
	Patient	-	Diabetic	-	Yes	-	Hyponymy	Patient	Diabetic	-
Lanes	Receptionist	Assistant	Assistant	Yes	-	-	Equivalent	Receptionist	Receptionist	Receptionist
	Doctor	Doctor	Medical Practitioner	Yes	-	-	Equivalent	Doctor	Doctor	Doctor
Messages	Appointment	-	Appt	Yes	-	No	Equivalent	Appointment	Appointment	-
	Invoice	Invoice	-	Yes	-	Yes	Different	Invoice_Patient	Invoice_ Supplier	-
Selected Elements	Invoice	Invoice	-	Yes	-	Yes	Different	Invoice_Patient	Invoice_ Supplier	-
	Appointment	-	Appt	-	-		Equivalent	Appointment	Appointment	-
	Patient	-	Patient File	Yes	-	No	Equivalent	Patient	-	Patient
Tasks	Display Appointment	-	Expose Appt	-	-	-	Equivalent	Display Appointment	-	Display Appointment
	Display Patient	-	Expose Patient File	-	-	-	Equivalent	Display Patient	-	Display Patient
	Search Patient	-	Search Patient File	-	-	-	Equivalent	Search Patient	-	Search Patient

We obtain the comparison table (TABLE II), after the elimination of duplicate elements. The first four columns present the elements of the models X, Y and Z to be compared, the fifth column presents whether or not the elements are syntactically or semantically equivalent. The sixth column shows whether or not the elements are hyponyms. The seventh column presents whether or not the elements are homonyms. The eighth column presents the final decision (equivalence, hyponymy or difference).

In order to choose the appropriate name of UML result elements (in the second subsystem of transformation), we base on the decision of column 8 to rename elements in the three last columns called RCM$_i$ (M$_i$ represents X, Y or Z), as following: (1) If the decision is "Equivalent", we rename the elements of model X, Y and or Z using the name of the first column. (2) If the decision is "Hyponymy", we keep the same name of the elements of the model. (3) If the decision is "Different", we use the name of the element and we add "_the name of the pool that is different to the pool of the other element".

2) Transformation of BPMN models into UML class diagram: In order to transform BPMN models into a UML class diagram we apply a transformation process (Fig. 9) based on the comparison table presented in section IV.C.1)c). This process is based on 18 transformation rules presented in detail in our previous work [29]. This process is presented in the form of a series of steps based on BPMN notation as follow (Fig. 9):

- Transformation of task: this sub-process is based on 12 rules related to tasks (considering their types and if they are linked to data or not) and call one other rule related to data. At the end of this step, the UML class diagram can be constituted of classes, attributes, operations, associations and multiplicities.
- Transformation of pool and/or lane: this sub-process generates classes, attributes, and aggregations after this step.
- Transformation of relationship between (pool or lane) and task: it generates associations and multiplicities.
- Transformation of message: it generates classes and attributes.
- Transformation of relationship between message and element (pool, task or event): it generates associations and multiplicities.

For those first fifth sub process, we:

- Identify an element "i" of a model M$_i$
- Apply the corresponding rule (according to the element identified). If the element belongs to the column M$_i$ in the comparison table, the name of the element (class or operation) that will be used corresponds to the RCM$_i$ column. In addition, a check is performed to determine whether an element (class, operation or association) of the class diagram has already been created by another rule. If it has:
    - o All the instructions associated with that rule are applied, except creation of the element.
    - o If an association already exists, such that the multiplicities are different, the existing association is kept, and the union of multiplicities is applied for each end of the association.

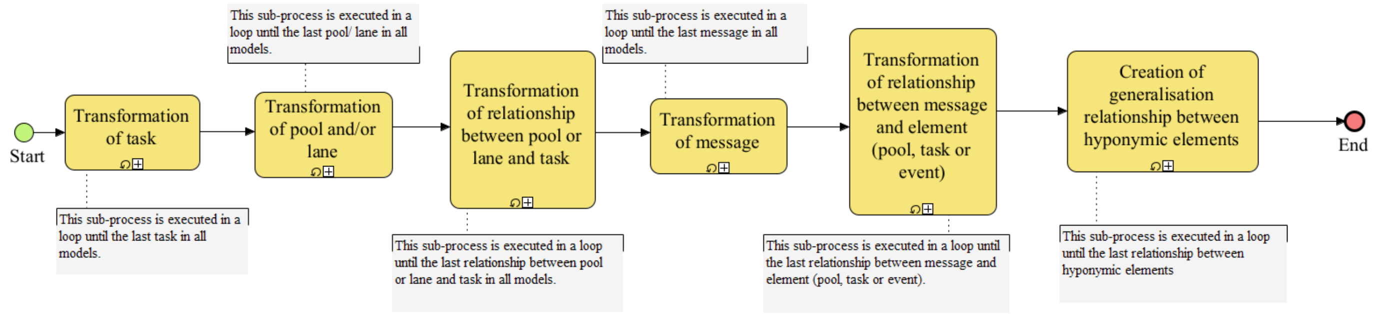

Start
Transformation of task
Transformation of pool and/or lane
Transformation of relationship between pool or lane and task
Transformation of message
Transformation of relationship between message and element (pool, task or event)
Creation of generalisation relationship between hyponymic elements
End
This sub-process is executed in a loop until the last task in all models.
This sub-process is executed in a loop until the last pool/ lane in all models.
This sub-process is executed in a loop until the last relationship between pool or lane and task in all models.
This sub-process is executed in a loop until the last message in all models.
This sub-process is executed in a loop until the last relationship between message and element (pool, task or event).
This sub-process is executed in a loop until the last relationship between hyponymic elements

Fig. 9. The Transformation Process.

- Creation of a generalisation relationship between hyponymic elements. In this sub-process, we:
  - o Identify hyponymic elements.
  - o Create a generalisation relationship between elements. When transforming an element "i" of model Mi, if the element belongs to the column Mi in the comparison table, the name of the element class that will be used corresponds to the RCMi column.

V. Case Study

In order to illustrate the application of the first step of our proposed approach, the three BPMN models represented in section IV.B, are transformed into a UML class diagram (Fig. 10) by applying our transformation process and rules and by referring to the comparison table. Resulted model represents the output of step 1 (M2' UML model).

We note that there is no redundancy in the classes. Indeed, for example, we find only the class "Medical Office" instead of having in addition the class "Office", only the class "Doctor, instead of having in addition the class "Medical Practitioner", only the class "Receptionist" instead of having in addition another class "Assistant", a single class "Appointment" instead of having in addition the class "Appt". Moreover, we don't find redundancies in class operations. In fact, for example, we find a single "displayAppointment()" operation, and not a second "exposeAppt()", we also found a single "displayPatient()" operation, and a single "searchPatient()" operation. Also, we don't find a loss information by finding the two classes "Invoice_Patient" and "Invoice_Supplier". Finally, we find a generalisation relationship between the "Diabetic" and "Patient" classes.

We present the resulted elements of the first step of transformation process (classes, attributes, associations, multiplicities) by black color. At the second step, the resulted elements (classes, attributes, aggregations, multiplicities) are illustrated using blue color. The resulted elements (associations and multiplicities) after the third step are mentioned using green color. After the fourth step, we didn't add new element because they were created in the previous steps. After the fifth step, the added or modified elements (associations, multiplicities) are illustrated by red color. Finally, we illustrate the resulted generalisation relationship, using yellow color.

To create the result UML model M2", we merge (by applying the fusion subsystem, already developed in our previous works), M2 model which can represent the existing software level in medical field and this M2' resulted model.

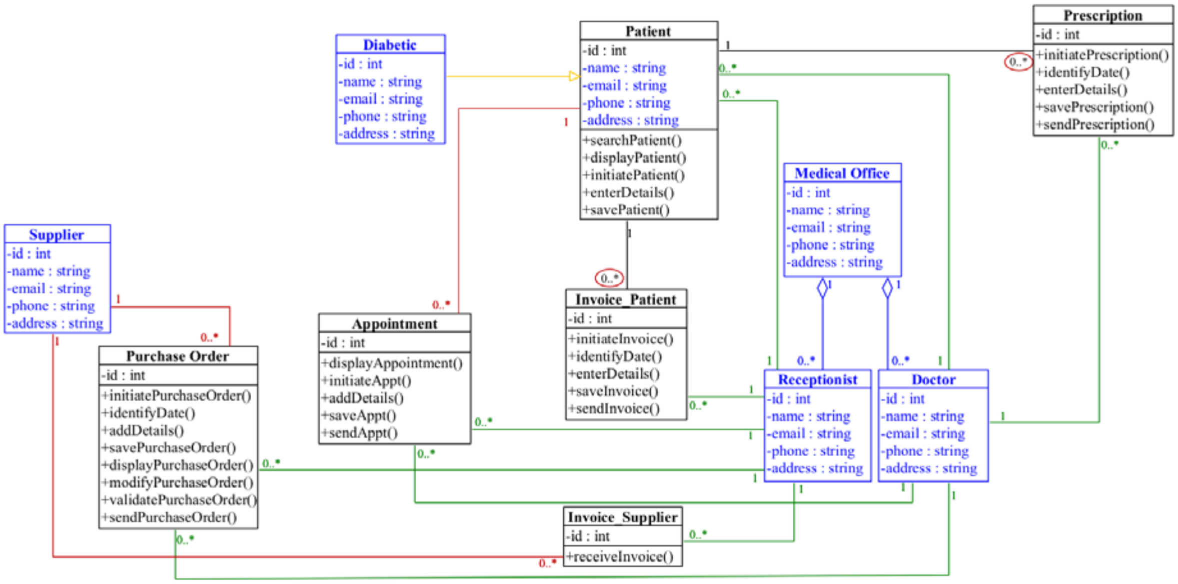


Fig. 10. Resulted Class Diagram UML Model M2'.

VI. Conclusion

In this paper, we propose a method for aligning software system level with business process level without conflicts, by considering a set of models at the source level.

We responded to the limitations of existing work. In fact, by applying the rules of syntactic and semantic comparison, we can detect real equivalences and real differences, and thus obtain a UML result model without conflicts. Indeed, by identifying two elements as semantically equivalent, we are left with only one element, and therefore do not have redundancy in UML diagram; additionally, identifying two elements as syntactically equivalent results in only one element, and thus no redundancy in result model. Identifying two elements as homonyms produces two result elements in UML model. Finally, identifying two elements as hyponyms produces a generalisation relationship between the two resulting elements in result model.

We have implemented a first version of our approach, which allows us to facilitate and automate the transformation phase. This was accomplished using the Java programming language in the Eclipse development environment. The tool takes a set of xpdl files as input. Each file is a representation of a BPMN model. An xmi file is generated from this input, which is later converted into a UML class diagram. It is tested through a case study in the field of telecommunications. Our intent is to update this module by considering syntactic and semantic aspects. We aim also proposing the alignment of BPMN models with UML model to realise a more complete alignment.

The syntactic and semantic comparison of BPMN models resulted in our work can be used in the first step of a research theme, which is the fusion of the BPMN models, in the context of organizations’ merge.

References

[1] T.Wasiuk, F.P.C.Lim, “Factors Influencing Business IT Alignment,” International Journal of Smart Business and Technology, vol.9, no.1, pp.1-12, Mar. 2021.

[2] H. Darii, J. Laval, V. Botta-Genoulaz, and V. Goepp, “Measurement of the business/IT alignment of information systems,” in ILS 2020-8th International Conference on Information Systems, Logistics and Supply Chain, pp. 228-235. 2020.

[3] P. Gajardo and L. P. Ariel, “The business-it alignment in the digital age,” In The 13th Mediterranean Conference on Information Systems (ITAIS & MCIS), Naples, Italy. 2019.

[4] M. Zhang, H. Chen, and A. Luo, “A systematic review of business-IT alignment research with enterprise architecture,” IEEE Access, vol. 6, pp. 18933–18944, 2018.

[5] A. Ullah and R. Lai, “A systematic review of business and information technology alignment,” ACM Trans. Manag. Inf. Syst., vol. 4, no. 1, pp. 1–30, 2013.

[6] O. M. G. UML, “OMG (2017) Unified Modeling Language®(OMG UML®) Version 2.5. 1 https://www. omg. org/spec,” 2017.

[7] Y. E. Chan, “Business Strategy, information system strategy, and strategic fit: Measurement and performance impacts,” p. 362, 1992.

[8] J. C. Henderson and H. Venkatraman, “Strategic alignment: Leveraging information technology for transforming organizations,” IBM Syst. J., vol. 38, no. 2.3, pp. 472–484, 1999.

[9] B. H. Reich and I. Benbasat, “Measuring the linkage between business and information technology objectives,” MIS Q., pp. 55–81, 1996.

[10] C. U. Ciborra, “Deconstructing the concept of strategic alignment,” Scand. J. Inf. Syst., vol. 9, no. 1, pp. 67–82, 1997.

[11] T. Smaczny, “Is an alignment between business and information technology the appropriate paradigm to manage IT in today’s organisations?,” Manag. Decis., 2001.

[12] J. Luftman, “Assessing business-IT allignment maturity,” in Strategies for information technology governance, Igi Global, pp. 99–128, 2004.

[13] D. W. Nickels, “IT-Business Alignment: what we know that we still don’t know,” in Proceedings of the 7th Annual Conference of the Southern Association for Information Systems, vol. 79, pp. 79-84. 2004.

[14] T. R. Gruber, “A translation approach to portable ontology specifications,” Knowl. Acquis., vol. 5, no. 2, pp. 199–220, 1993.

[15] P. Zweigenbaum, “MENELAS: an access system for medical records using natural language,” Comput. Methods Programs Biomed., vol. 45, no. 1–2, pp. 117–120, 1994.

[16] T. R. Gruber, “Toward principles for the design of ontologies used for knowledge sharing?,” Int. J. Hum. Comput. Stud., vol. 43, no. 5–6, pp. 907–928, 1995.

[17] M. Habba, M. Fredj, and S. B. Chaouni, “Towards an operational alignment approach for organizations,” in Proceedings of the 9th International Conference on Information Management and Engineering, pp. 29–34, 2017.

[18] M. Habba, M. Fredj, and S. Benabdellah Chaouni, “Alignment between Business Requirement, Business Process, and Software System: A Systematic Literature Review,” J. Eng. (United Kingdom), vol. 2019, 2019, doi: 10.1155/2019/6918105.

[19] M. F. Amr, N. Benmoussa, K. Mansouri, and M. Qbadou, “Transformation of the CIM Model into A PIM Model According to The MDA Approach for Application Interoperability: Case of the" COVID-19 Patient Management" Business Process,” iJOE, vol. 17, no. 05, p. 49, 2021.

[20] D. Brdjanin, G. Banjac, and S. Maric, “Automated synthesis of initial conceptual database model based on collaborative business process model,” in International Conference on ICT Innovations, pp. 145–156, 2014.

[21] D. Brdjanin, A. Vukotic, G. Banjac, D. Banjac, and S. Maric, “Automatic Derivation of Conceptual Database Model from a Set of Business Process Models,” in 2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), pp. 1–8, 2020.

[22] W. Khlif, N. Elleuch, E. Alotabi, and H. Ben-Abdallah, “Designing BP-IS Aligned Models: An MDA-based Transformation Methodology,” pp. 258– 266, 2018.

[23] Y. Rhazali, Y. Hadi, and A. Mouloudi, “A methodology of model transformation in MDA: From CIM to PIM,” Int. Rev. Comput. Softw., vol. 10, no. 12, pp. 1186–1201, 2015, doi: 10.15866/irecos.v10i12.8088.

[24] E. F. Cruz, R. J. Machado, and M. Y. Santos, “From business process modeling to data model: A systematic approach,” in 2012 Eighth International Conference on the Quality of Information and Communications Technology, pp. 205–210, 2012.

[25] E. F. Cruz, R. J. Machado, and M. Y. Santos, “Deriving a Data Model from a Set of Interrelated Business Process Models.,” in ICEIS (2), pp. 49–59, 2015.

[26] A. Kriouile, N. Addamssiri, T. Gadi, and Y. Balouki, “Getting the static model of PIM from the CIM,” in 2014 Third IEEE International Colloquium in Information Science and Technology (CIST), pp. 168–173, 2014.

[27] B. Bousetta, O. El Beggar, and T. Gadi, “A methodology for CIM modelling and its transformation to PIM,” Journal of Information Engineering and Applications, vol. 3, no. 2, pp. 1–21, 2013.

[28] B. P. M. OMG, “Notation (BPMN) Version 2.0. 2, Object Management Group, 2013,” 2016.

[29] M. Habba, S. B. Chaouni, and M. Fredj, “Aligning Software System Level with Business Process Level through Model-Driven Architecture,” International Journal of Advanced Computer Science and Applications, vol. 12, no. 10, pp. 174–183, 2021, doi: 10.14569/IJACSA.2021.0121020.

[30] S. B. Chaouni, M. Fredj, and S. Mouline, “Metamodels for models complete integration,” in 2011 IEEE International Conference on Information Reuse & Integration, 2011, pp. 496–497.

[31] Ontology Definition Metamodel, OMG Adopted Specification, OMG Document Number: ptc/2008-09-07, 2008.

Multi-Task Reinforcement Meta-Learning in Neural Networks

Ghazi Shakah
Software Engineering Department, Faculty of Information and Technology
Ajloun National University, P.O. Box 43, Ajloun 26810, Jordan

***Abstract*—Artificial Neural Networks (ANN) is one of the main and widespread tools for creating intelligent systems. And, they are actively used for data analysis in many areas such as robotics, computer vision, natural language processing, etc. The learning process of ANN is one of the most labor-intensive stages in ANN. There are many different modifications of ANNs and methods for their training. Currently, deep neural networks are becoming one of the most popular methods of machine learning due to their effectiveness in areas such as speech recognition, medical informatics, computer vision, etc. It is known that ANN training depends on the type of input data. In this paper, reinforcement learning is considered, as popular method used in cases where information is reinforced by signals from the external environment with which the model interacts. The purpose of this paper is to develop a reinforcement meta-learning algorithm that would be efficient in terms of quality and speed of learning. However, despite the significant scientific progress in deep learning, existing algorithms are not efficient enough to solve problems in the real world. In addition, such algorithms require a significant amount of learning time, which complicates the development process. To solve these problems, the use of meta-learning or "learning to learn" algorithms has recently been especially relevant. The paper proposes an approach to reinforcement meta-learning using a multitasking weight optimizer. experimentally shown that the proposed approach is more efficient than the known MAML (Model-Agnostic Meta-Learning) algorithm. The proposed MAML SPSA-Track method shows an improvement in efficiency by an average of 4%, and MAML SPSA-Delta by 8%, respectively. Moreover, the last algorithm spends on average 2 times less time on push-v2 and pick-place-v2 tasks.**

***Keywords*—Multitasking; meta-learning; reinforcement learning; neural networks; optimization**

I. Introduction

Humans have an innate ability to learn new skills quickly and easily. For example, we can look at one instance of a knife and distinguish all knives from other cutlery such as spoons and forks. Our ability to learn new skills and quickly adapt to a new environment based on a small number of examples is not just limited to identifying new objects, learning a new language, or figuring out how to use a new tool; our possibilities are much more diverse. [1], [2]. In contrast, machines—specifically, deep reinforcement learning algorithms—generally learn quite differently [3]. They require very large amounts of data and computational resources to achieve acceptable performance. The reason why people can learn quickly and adapt to a new environment is that they use the knowledge gained from previous experience to solve new problems. Similarly, meta-learning uses little experience gained from data to solve new problems quickly and efficiently. Through this method, it is possible to significantly speed up the training of neural networks with reinforcement significantly. Neural networks with reinforcement require quite large amounts of training data and computational resources. Creating such datasets is costly, especially when you need to involve a domain expert. While pre-training is useful, these approaches become less efficient for domain-specific problems, which it still requires large amounts of task-specific labeled data to achieve good performance. In addition, some existing problems are characterized by a wide and unbalanced distribution of data, which can make it difficult to collect training examples [4]. On the other hand, it is possible to use a pre-trained network from another task and then finish training it on the current small training set. However, depending on the specifics of the problem, this is not always possible, especially if the task on which the neural network was trained is significantly different. It is important to note that the ability to quickly learn new tasks during model inference is something traditional machine learning approaches do not attempt. This is what makes meta-learning especially attractive. Meta-learning is particularly interesting and can be used for the following reasons [5].

- The ability to learn from just a few examples.
- Quick adaptation to new tasks.
- The ability to create more versatile systems.

Meta-learning is especially successful in situations where a large amount of data is required; for example, robots are tasked with learning new skills in the real world and often encounter new environments [6].

Finally, the task was formally set to develop a meta-learning algorithm with reinforcement of a machine learning model that would be efficient in terms of learning.

II. Literature Review

Meta-learning tries to gain general knowledge about the target area by learning the set of tasks belonging to it [7]. The idea of meta-learning is to train the model by showing it only a few examples for each class and [8] then test it on new examples from the same classes that were taken from the original dataset.

The author in [9] proposed a formal description of the few-shot learning task as meta-learning. The data set of each class is randomly divided into a support set and a query set. The support set consists of labeled examples that are used to predict classes of untagged examples from the query set. The set of classes at the training stage does not intersect with the set of classes at the testing stage.

The author in [10] introduced a multitasking loss function based on maximizing the Gaussian likelihood with a task-dependent uncertainty. The proposed single model for all tasks outperformed separate models for each task.

Subsequently, a multitasking approach was applied in [11] to solve the problem of character recognition from a small number of examples, which led to an improvement in the overall recognition efficiency compared to the base model.

The author in [12] established a close relationship between the optimization problems of multitask learning and optimization-based meta-learning. Different from existing works, this paper focuses on improving the meta-learning stage. Only inductive methods that use a meta-learning process without prior training are considered. To do this, MAML was chosen as an example of the use of optimization-based learning, since the works describing this method are among the most cited in this field.

III. Material and Method

A. The Task of Reinforcement Meta-Learning

Reinforcement learning (RL) is one of the methods of machine learning, the purpose of which is to find an optimal strategy for the behavior of the model in the environment and maximize the reward received from the environment throughout the entire time the model interacts with the environment. The main concepts in RL are the agent and the environment: The environment represents the world in which the agent lives and interacts. In Fig. 1, at each interaction step, the agent observes (perhaps only partially) the state of the environment [3]. While the agent then decides what action to take. The environment changes when the agent acts on it, but it can also change by itself. The agent also receives a reward from the environment, a number that tells him how good or bad the current state of the world is. Accordingly, the agent's goal is to maximize his total reward, called profitability. Reinforcement learning methods are approaches by which an agent can learn the desired behavior to achieve a goal [13].

Reinforcement meta-learning is meta-learning in the field of reinforcement learning. Usually, training and test problems are different, but they are taken from the same family of problems.

Let's we have a distribution of tasks, each of which is The Markov Decision Process (MDP).

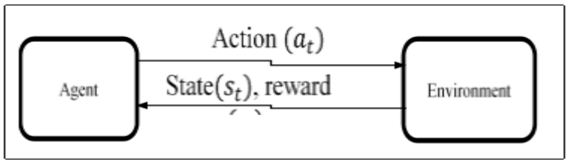


Fig. 1. The Cycle of Interaction between the Agent and the Environment [3].

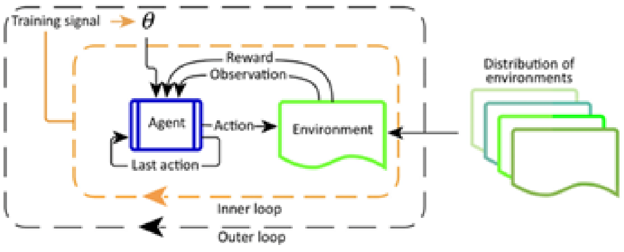


Fig. 2. A Meta-RL representation Containing Two Optimization Loops [4].

$M_i \in M$, where M_i is defined by the set $\langle S, A, P_i, R_i \rangle$. In Fig. 2, at each iteration of the external cycle, a new environment is selected and the parameters that determine the behavior of the agent are adjusted using the metal earning algorithm. In the inner loop, the agent interacts with the environment and maximizes the reward using a reinforcement learning algorithm [4], [14]. Note that the general state S and action space A, so the stochastic policy is:

$\pi_\theta: \mathrm{S} \times \mathrm{A} \rightarrow \mathrm{R}_+$

Will receive input data that is compatible with different tasks. Test items are selected from the same or slightly modified distribution M. In general, reinforcement meta learning is very similar to regular reinforcement learning, except that the last reward r_{t-1} and the last action $a_{t-1}$are also included in the observation in addition to the current state. s_t:

- In reinforcement learning $\pi_\theta(s_t) \rightarrow a$
- In Reinforcement Meta-learning $\pi_\theta(a_{t-1}, r_{t-1}, s_t) \rightarrow a$

This is done so that the policy will learn the changes between states, rewards and actions in the current MDP and can adjust its strategy accordingly. This is done so that the policy can assimilate the changes between states, rewards and actions in the current MDP and can adjust its strategy accordingly.

IV. Development of the Training Method

A. MAML Meta-Learning Algorithm

Reinforcement meta-learning uses two optimization loops: external and internal. During the outer loop, a meta-learning algorithm is applied. It is important to note that MAML is compatible with any model that can be trained using gradient descent, which is its main advantage. There aren't any restrictions on the loss function. The algorithm is applicable to such a wide range of problems as regression, classification, and reinforcement learning. MAML does not change the structure of the learning model, but only changes the network parameters in such a way that a small number of gradient descent steps are required on a small training dataset of a new problem to obtain a good generalization ability on this problem [15]. However, this algorithm requires taking second-order derivatives, which is the main disadvantage of this algorithm.

B. Method Formulation of the Problem

Despite the fact that there are a number of algorithms that do a good job of this kind of task, the speed of learning these algorithms, even on the most productive equipment, takes an extremely long time. Moreover, due to the dynamism and variety of tasks in the real world, the quality of the solution on new test problems can be much worse than the quality of training ones. So, the goal of this thesis is to develop a reinforcement meta-learning algorithm that would be effective in terms of quality and speed of learning. Also, during the execution of the work, the following tasks were set:

- Modify the basic MAML (Model-Agnostic Meta-Learning). An algorithm based on a multitasking approach.
- The tasks of deep learning, reinforcement learning, meta-learning, multitasking learning, and reinforcement meta-learning have been described.
- The main approaches to meta-learning were analyzed - based on models, metrics and optimization, and modern meta-learning algorithms for each approach.
- Compare the efficiency of the modified and unmodified algorithms.
- The common problem of low efficiency for these methods was analyzed in terms of both data and time resources spent on training the model.
- Formally, the task was to develop a meta-learning algorithm with reinforcement of a machine learning model that would be effective in terms of quality and speed of learning.

V. Software Implementation of the Console Application

A. Architecture and Composition

The console application was developed in the popular Python programming language for further experiments. Fig. 3 describes the standard process for developing a strategy model, along with the important modules associated with solving the problem.

Meta-World's ML1 environment was used as an environment for the agent. To implement the neural network of the agent, the torch.nn module of the well-known PyTorch framework was used. [16].

To realize the basic MAML algorithm and the proposed SPSA-Delta and SPSA-Track algorithms, the modules were optim and torch. Autograd was used, which presents various standard optimization algorithms for training neural networks. The training results were written to the hard disk using the torch.utils.tensorboard module. PyTorch is one of the most popular open-source machine learning frameworks in the Python programming language.

The main PyTorch modules that are used in the software implementation are: torch.nn, torch.optim, torch.autograd, torch.utils.tensorboard. The torch.nn module defines computational graphs and works with gradients, which makes it easy to build neural networks. The following module torch.optim introduces various optimization algorithms for training neural networks. The torch.autograd module implements the automatic differentiation method. The torch.utils.tensorboard module helps to save and visualize the results.

The first step in any deep learning project involves loading and processing training data. Reinforcement learning of a model consists of its interaction with the simulated environment. Meta-World allows you to design environments according to the Env interface of the Gym framework. First, we need to create the desired test, and then an instance of the environment. A task is assigned to the environment using the set_task() method from the corresponding already defined training and test tasks of the created test. In the current project, a function was described that returns an instance of the environment given the benchmark test and the task name task name.

The process of creating the environment of the Meta-World framework for the subsequent training of the model is shown in Fig. 4. The agent's interaction with the environment is implemented through the environment's step() method. For the convenience of interacting with the environment, the Runner class was described, an instance of which receives from the model the action to be performed, passes it to the simulated environment and receives from it information about the current state, the value of the reward, success rate and other metadata. Then, an instance of the ReplayMemory class collects all the received information about states, actions, etc. into the corresponding tensors in order to further transfer it to the main MAML algorithm for processing. The strategy is presented as a neural network. To create it, PyTorch uses the corresponding torch.nn module. It provides the implementation of all commonly used neural network components such as fully connected and convolutional layers, activation layers and associated loss functions.

The neural network representing the main strategy consists of one hidden layer with a size of 128 neurons and an activation function nn.Tanh() between the layers Fig. 5.

The input of the neural network is a vector of length 39 about the state of the ML1 Meta-World test environment, at the output the neural network gives a vector of length 4 about the next action by the agent in the environment.

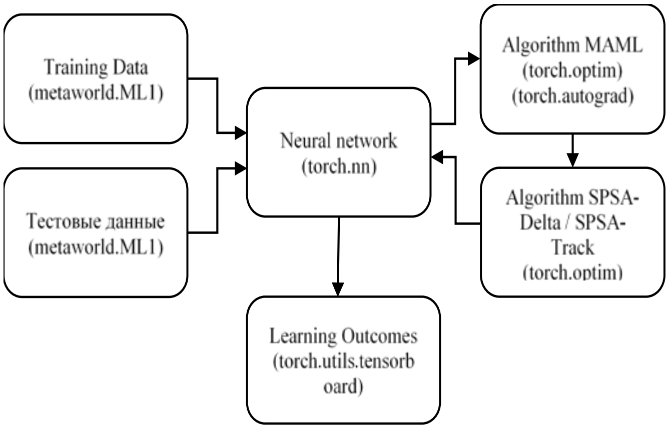
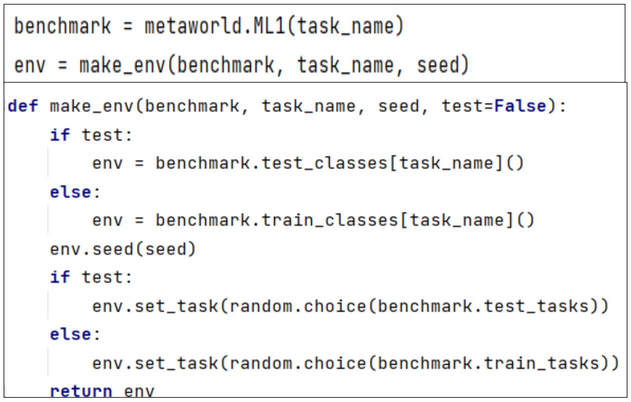

Training Data (metaworld.ML1)
Тестовые данные (metaworld.ML1)
Neural network (torch.nn)
Algorithm MAML (torch.optim) (torch.autograd)
Algorithm SPSA-Delta / SPSA-Track (torch.optim)
Learning Outcomes (torch.utils.tensorboard)

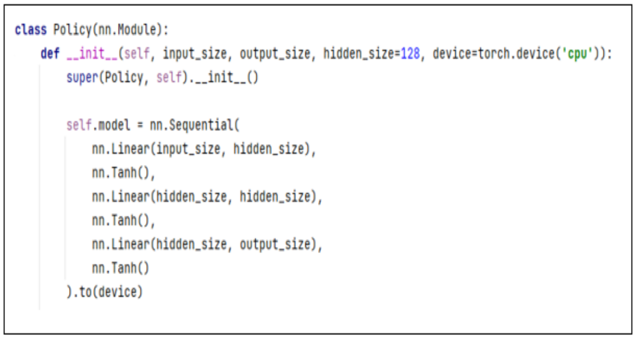
Fig. 3. Diagram of the Development Process of a Neural Network Model.



```
benchmark = metaworld.ML1(task_name)

env = make_env(benchmark, task_name, seed)
```

```
def make_env(benchmark, task_name, seed, test=False):
    if test:
        env = benchmark.test_classes[task_name]()
    else:
        env = benchmark.train_classes[task_name]()
    env.seed(seed)
    if test:
        env.set_task(random.choice(benchmark.test_tasks))
    else:
        env.set_task(random.choice(benchmark.train_tasks))
    return env
```

Fig. 4. The Initialization of the Meta-World Environment.



```
class Policy(nn.Module):
    def __init__(self, input_size, output_size, hidden_size=128, device=torch.device('cpu')):
        super(Policy, self).__init__()

        self.model = nn.Sequential(
            nn.Linear(input_size, hidden_size),
            nn.Tanh(),
            nn.Linear(hidden_size, hidden_size),
            nn.Tanh(),
            nn.Linear(hidden_size, output_size),
            nn.Tanh()
        ).to(device)
```

Fig. 5. The Strategy Neural Network Initialization.

For each iteration of the meta-learning stage 10 training tasks were selected to obtain the needed data. The number of adaptation steps of the MAML algorithm was equal to 1, i.e. during testing, adaptation to a new problem occurred in one step of the gradient descent algorithm. The maximum length for a single task in the environment was 500. The maximum length for a single task in the environment was 500. The reward discount factor was 0.99.

During the adaptation phase, the weights of the neural network were changed with a learning rate factor of 1×10^{-4}, during the meta-learning phase 1×10^{-3}. The optimization method TRPO was used as a meta-optimizer with a maximum number of steps for linear search of 15 and a step of 5×10^{-3}.

The neural network was trained for 600 epochs for environment with reach-v2, pick-place-v2 tasks and 1200 epochs for push-v2. Also, the multitasking weight optimizer only started optimization after 50 epochs so that the model had time to adjust the initially randomized weights according to the task and environment.

B. General Description of the MetaWorld Environment

Meta World is an open-source simulated test for reinforcement meta-learning and multitasking learning, consisting of 50 different environments with robotic manipulations [17]. Task T in Meta World is defined as a set consisting of a reward function, an object's starting position, and its target position. Metal-earning makes two important assumptions:

1) Meta-training and meta-testing tasks have a common distribution $p(T)$.

2) The task distribution $p(T)$, has a general structure that can be used to effectively adapt to new tasks.

If $p(T)$, is defined as a family of variations within a specific problem, as in previous works [6], [10], then it is unreasonable to hope for a generalization to completely new problems. For example, an agent has little chance of being able to quickly learn to open a door without ever hitting a door before if it has only been trained on a set of uniform and narrow tasks during meta-learning. Thus, in order to allow reinforcement meta-learning methods to adapt to completely new tasks, a sufficiently large set of tasks is needed, where continuous changes in parameters cannot be used to describe the differences between tasks. In Meta World, all tasks are performed by a robotic arm. The action space is a set of two elements, consisting of changing the 3D space of the gripper.

Actions in this space range from -1 to 1. The observation space is represented as a set of 6 3D Cartesian positions of the gripper, a normalized measurement of how open the grip is, the 3D position of the first object, the quaternion of the first object, the 3D position of the second object, the quaternion of the second object, all previous measurements in the environment, and finally 3D position of the target. If there is no second object or the target is not supposed to be included in the observation, then the values corresponding to them are set to zero. The state space is always 39-dimensional. The reward functions for all tasks have the same value, which is in the range from 0 to 10, where 10 always corresponds to the fact that the task was solved. It should be noted that all tasks were implemented using the MuJoCo physics engine [18], [19], which allows to simulate the physical laws of the real world quickly and efficiently. The Multi-world interface [20] and interfaces of the popular OpenAI Gym environment [21] were taken as the basis, which makes this framework quite easy to use.

VI. Experimental Methodology

As an applied task for solving and evaluating the efficiency of the developed algorithms, we consider the ML1 test in the Meta-World environment. ML1 is aimed at evaluating the adaptation of the algorithm in several steps to change the goal within the same task. ML1 uses separate Meta-World tasks, where training tasks correspond to 50 random initial positions of objects and targets, and testing tasks correspond to 50 held positions. Algorithms are evaluated on three tasks from Meta-World:

- reach-v2.
- push-v2.
- pick-place-v2.

where either the position to be reached or the target position of the object varies. Target positions are not specified in world states, which forces reinforcement meta-learning algorithms to adapt to the target through trial and error.

In the reach-v2 task, the robotic arm needs to reach a target position, which is given randomly. The next push-v2 challenge is to push the puck towards the net. In the pick-place-v2 problem, it’s important to take and place the puck in the goal. The positions of the puck and the goal in the tasks are set randomly.

Since the reward values do not directly indicate the success of the chosen policy, Meta-World defines an interpretable success metric for each task, which is used as the main criterion. Since all tasks involve the manipulation of one or more objects in the target configuration, this success metric is usually based on the distance between the task-relevant object and its final target position, i.e.,$||o - j|| < \varepsilon$, where ε is the threshold, for example, 5 cm. The software implementation was carried out in the Python programming language version 3.8. The environment was modeled using the Meta-World framework version 2.0 together with the OpenAI Gym framework version 0.19. As noted earlier, MetaWorld contains ready-made implementations of various environments for meta-reinforcement learning and agent testing. To build models of neural networks, the PyTorch framework version 1.10 was used, which contains implementations of various layers and algorithms for optimizing neural networks.

VII. Analysis of Results

Fig. 6 shows the maximum success rate, averaged over 5 runs, in the ML1 Meta-World test environment. Based on the results obtained, all 3 algorithms do an excellent job of solving the reach-v2 problem both at the training and testing stages. On more complex push-v2 and pick-place-v2 tasks, the MAML SPSA-Delta algorithm is the most efficient among all those considered. The improvement relative to the basic algorithm was 17% at the training stage and 21% at the testing stage on the push-v2 task, 8% and 3% on the pick-place-v2 task, respectively. However, on the pick-place-v2 problem, the ability of the MAML SPSA-Delta method to generalize is not much higher than the MAML SPSA-Track algorithm (60% for MAML SPSA-Delta and 59% for MAML SPSA-Track). The following figures show examples of the moving average success rate for the reach-v2 task over 6×10^7steps in a test environment (Fig. 6 is the training phase, Fig. 7 is the testing phase). For all constructed charts, the moving average coefficient is 0.8. As can be seen from the graph, the modified MAML algorithms solve the problem no worse than the original MAML algorithm environment (Fig. 7 is the training phase; Fig. 8 is the testing phase). For all constructed charts, the moving average coefficient is 0.8.

As can be seen from the graph, the modified MAML algorithms solve the problem better than the original MAML algorithm. Now let's look at the moving average success rate plots for the following push-v2 task over 2×10^8steps in a test environment (Fig. 9 is the training phase, Fig. 10 is the testing phase). Compared to other tasks, the SPSA-Delta MAML algorithm took significantly longer to adapt to the environment and overtake the basic MAML method. The SPSA-Track algorithm also shows good results, but they do not differ significantly from the results of the unmodified MAML algorithm. Finally, let's analyze the constructed plots of the moving average success rate for the last pick-place-v2 task over 6×10^7 steps in the test environment (Fig. 11 is the training phase, Fig. 12 is the testing phase). It follows from the graph that both MAML algorithms with modifications are significantly more efficient at the testing stage than the original MAML algorithm.

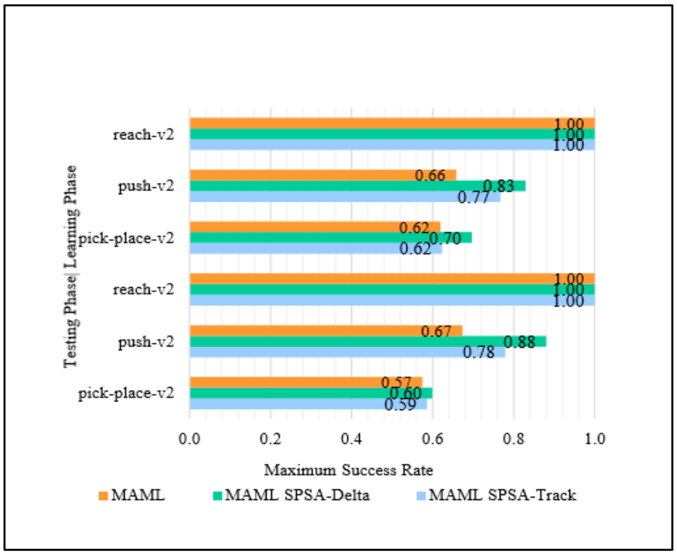


Fig. 6. The Maximum Success rate of Algorithms in the ML1 Meta-World Test Environment.

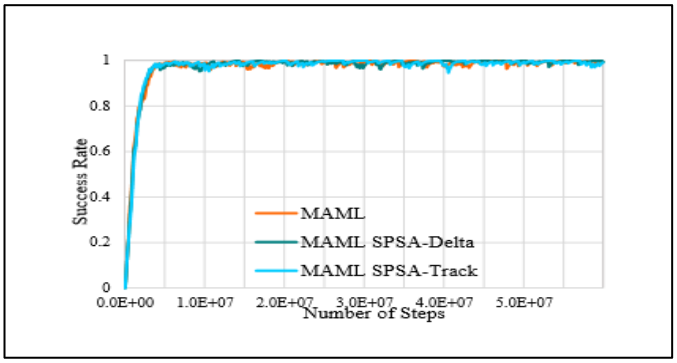


Fig. 7. Average Success Rate of Algorithms on the Reach-v2 Problem at the Training Stage.

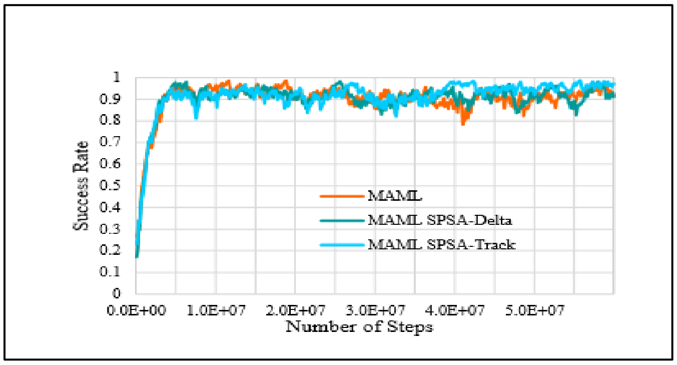


Fig. 8. Success Rate of Algorithms on the Reach-v2 Problem at the Testing Stage.

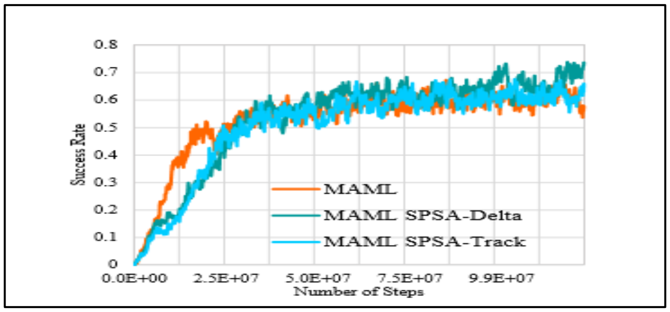


Fig. 9. Success Rate of Algorithms on the Push-v2 Problem at the Training Stage.

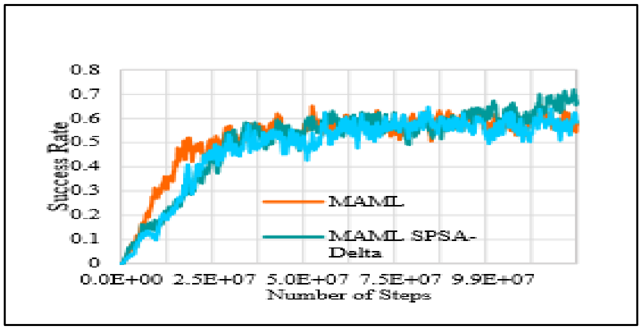


Fig. 10. Success Rate of Algorithms on the Push-v2 Task at the Testing Stage.

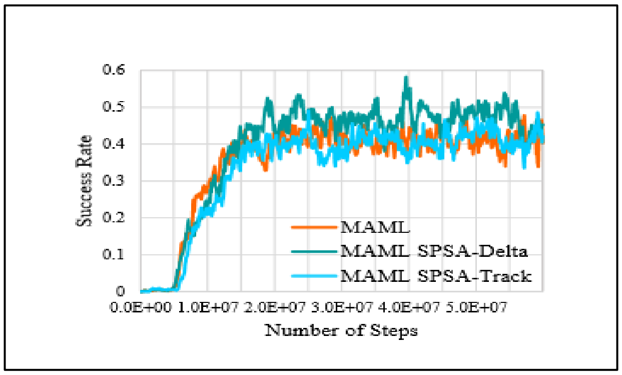


Fig. 11. Success Rate of Algorithms on the Pick-place- v2 Problem at the Training Stage.

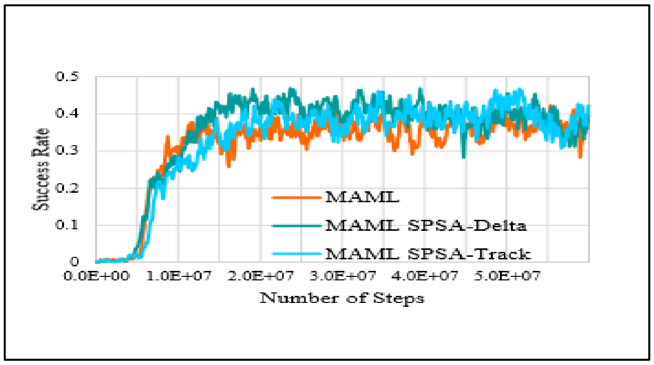


Fig. 12. Success Rate of Algorithms on the Pick-place-v2 Task at the Testing Stage.

Moreover, the MAML SPSA-Delta and MAML with SPSA-Track algorithms achieve the same high success rate, but the first algorithm learns 2 times faster than the second.

Consider the average maximum success rate achieved by each algorithm in the ML1 Meta-World test environment, information about which is shown in Table I.

Based on the results of experiments in the ML1 MetaWorld test environment with three different methods of deep reinforcement meta-learning: the original MAML algorithm, MAML SPSA-Delta, MAML SPSA-Track, the proposed MAML SPSA-Track method shows an average efficiency improvement of 4%, and MAML SPSA-Track Delta by 8%, respectively. Moreover, the latter spends on average 2 times less time for training on push-v2 and pick-place-v2 tasks. According to the obtained results, it is safe to say that the use of a multitasking loss function and its stochastic approximation with simultaneous perturbation can significantly improve the efficiency of deep reinforcement learning algorithms.

TABLE I. MAXIMUM SUCCESS RATE AVERAGED OVER ALL TASKS IN THE ML1 META-WORLD TEST ENVIRONMENT.

Algorithms	**Learning Phase**	**Testing Phase**
MAML	**76%**	75%
MAML SPSA-Delta	84%	83%
MAML SPSA-Track	80%	79%

VIII. CONCLUSION

Based on the tasks of deep learning, reinforcement learning, meta-learning, meta-learning with reinforcement and multitasking learning and their relevance are described.

After making comparison of the basic MAML algorithm with the proposed MAML SPSA-Delta and MAML SPSA-Track by conducting computational experiments to train the agent on reach-v2, push-v2, pick-place-v2 tasks in the ML1 Meta World test environment, it was concluded that the MAML SPS-Track algorithm is on average 4% more efficient compared to the original MAML method, and the MAML SPSA-Delta algorithm is 8% more efficient. Moreover, the last algorithm spends on average 2 times less time on push-v2 and pick-place-v2 tasks.

REFERENCES

[1] Wang, Y. X. Ramanan, D. and Hebert M. " Learning to Model the Tail". Advances in Neural Information Processing Systems 30, 2017, pp.58-69.

[2] Zidong Zhang, Dongxia Zhang, and Robert C. Qiu," Deep reinforcement learning for power system applications: an overview".,csee journal of power and energy systems, vol. 6, no. 1, march 2020,pp.213-225.

[3] Martín H., J. A., de Lope, J., and Maravall, D. "The kNN-TD Reinforcement LearningAlgorithm", Lecture Notes in Computer Science,2009, pp.305-314.

[4] Ezzeldin, T. and Kassis, A." Beyond Explanations: Recourse via Actionable Interpretability – Extended",Research gate, 2020, pp.1-17.

[5] L. David, A. Michael, and B. Richard. " Dynamic core competences through meta-learning and strategic context ",Elsevier, Journal of Management, Volume 22, Issue 4, 1996, PP. 549-569.

[6] G. Shakah. " The Devices of the Internet of Things Based on the Recognition of Handwriting Words with Mobile Assisted ", International Journal of Interactive Mobile Technologies (iJIM),Vol.14, No. 4,2020, pp.74-85.

[7] O. Aissam, E,Hamza, and J, Philippe Leroy. " Dynamic Access Control Policy based on Blockchain and Machine Learning for the Internet of Things", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 8, No.7, 2017, pp.417-424.

[8] R. Vilalta, and D.Youssef ." A perspective view and survey of meta-learning. Artificial Intelligence Review", Vol.18, No2, 2002,pp.77–95.

[9] O. Vinyals, C. Blundell, T. Lillicrap, k. kavukcuoglu, and D.Wierstra. " Matching networks for one shot learning. Advances in neural information", processing systems, Vol,29, 2016, pp.3630–3638.

[10] A. Kendall, Y. Gal, and R. Cipolla." Multi-task learning using uncertainty to weigh losses for scene geometry and semantics", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7482–7491.

[11] A. Boiarov, O. Granichin, and O Granichina. " Simultaneous perturbation stochastic approximation for few-shot learning " , IEEE, European Control Conference (ECC), 2020, pp. 350–355.

[12] H. Wang, H. Zhao, and B. Li. "Bridging multi-task learning and meta-learning: Towards efficient training and effective adaptation". Vol,139. In International Conference on Machine Learning. PMLR, 2021, pp. 10991-11002.

[13] H. Kono, Y. Murata G, A.Kamimura, K. Tomita, and T,Suzuki. "Transfer Learning Method Using Ontology for Heterogeneous Multi-agent Reinforcement Learning", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 5, No. 10, 2014,pp.156-164.

[14] M. Botvinick M, S. Ritter, X.Wang, Kurth-Nelson Z. Blundell.,and D. Hassabis . "Reinforcement Learning, Fast and Slow" , Trends in Cognitive Sciences, Vol. 23, No, 5 , 2019, pp. 26-41.

[15] A. Nichol, J. Achiam, and J. Schulman. "On First-Order Meta-Learning Algorithms", NIPS, 2018, pp.55-69.

[16] S. Jadon, and A. Garg. "Hands-On One-shot Learning with Python" Packet Publishing, 2020.

[17] T. Yu, D. Quillen, Z. He, R. Julian, K. Hausman, C. Finn, and S. Levine, "Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement Learning", Proceedings of the Conference on Robot Learning, PMLR 100, 2020, pp.1094-1100.

[18] E. Todorov, T. Erez, Y. Tassa. And A. Mujoco." physics engine for model-based control", In International Conference on Intelligent Robots and Systems, IEEE, October 2012.

[19] A. Al-Oqaily, and G .Shakah, "solving Non-linear Optimization Problems Using Parallel Genetic Algorithm", International Conference on Computer Science and Information Technology (CSIT), IEEE, 2018.pp.103-106.

[20] A. V. Nair, V. Ong. M. Dalal. S. Bahl, S. Lin, and S. Levine. "Visual reinforcement learning with imagined goals", Advances in Neural Information Processing Systems 31, 2018.

[21] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. Openai gym. arXiv:1606.01540, 2016.

Data Fusion Model of Road Sensors based IoT Feature Clustering

Hua Yang*
Civil Engineering Department
Guangxi Vocational Normal University, Nanning, China

***Abstract*—The collection of traffic data can play a role in analyzing and predicting highway design, planning, and real-time traffic management. The accuracy requirements for road dynamic data collection are low, and the accuracy is usually 3%-5%. However, it is required that vehicles can pass at high speed and obtain traffic information such as vehicle classification and vehicle speed. The prerequisite for the application of Internet of Things (IoT) technology to road information monitoring lies in the research and development of sensor technology in the perception layer and communication technology in the network layer, so that can obtain a large amount of perception data to serve the development and application of algorithms. To achieve the goal of low-cost and long-term monitoring of comprehensive traffic information and road service status information, this paper constructs a road vibration monitoring system, carries out road vibration monitoring under complex road environments, and proposes a traffic information monitoring method driven by road vibration data. By deploying the pavement vibration monitoring system in the actual road, the original signal of pavement vibration under the action of vehicle moving load is obtained. Through smooth processing and eigenvalue extraction, the monitoring of vehicle speed, wheelbase driving direction, vehicle load position and traffic flow is realized. The experimental results prove that the analysis of the road dynamic response under working conditions, as well as smoothing processing and eigenvalue extraction, the numerical modeling method in this paper realizes the monitoring of the position of the vehicle load and the traffic flow. The calculation error of vehicle speed and wheelbase is within ±4%, which is helpful to find the characteristic index of road vibration signal for evaluating road service status, and provides a reference for the application of road vibration response in road damage early warning and scientific maintenance.**

Keywords*—Traffic data; Internet of Things (IoT) technology; perception sensors; vibration monitoring; k-means++ algorithm

I. Introduction

The road is the infrastructure for all kinds of trackless vehicles and pedestrians. Road transportation can realize the reasonable allocation of production materials, integrate various resources, promote the flow of materials, talents, information and funds between regions, and ultimately promote economic development [1]. Due to factors such as heavy traffic, heavy-axle trucks, high tire pressures, and harsh weather conditions, many roads have developed fatigue cracking, loose materials, rutting and other diseases, which are likely to cause traffic congestion and accidents. It also severely shortens the service life of the road and reduces the driving efficiency [2]. How to extend the service life of roads and improve the efficiency of transportation under the actual driving load and the external environment has always been a hot issue of concern to the academic and industrial circles at home and abroad. Therefore, it is necessary to establish a comprehensive, intelligent and long-lasting road dynamic response monitoring system to understand the road dynamic response under vehicle load in time. Furthermore, real-time monitoring of road service status and road traffic conditions is carried out to realize scientific road maintenance and intelligent traffic control [3].

Many road information monitoring projects have been carried out at home and abroad. For example, more than 400 sensors are embedded in the test section of the intelligent highway in Virginia, USA to collect data such as stress and strain, temperature and humidity, and traffic volume [4]. Literature [5] develops a stress-based analytical model of pavement performance by burying strain gauges inside the pavement, and points out that the impact of temperature and water on pavement performance should be considered. Literature [6] acquires the lateral and longitudinal strain information of the road surface under different vehicle loads under various working conditions by embedding stress and strain sensors, and establishes the relationship function between the pavement structure modulus and the stress and strain. These works have made many contributions to road dynamic response monitoring. However, because the road structure is different from other building structures, its distribution area is wide, and it is affected by vehicle load and environmental factors for a long time. These will lead to some shortcomings in the existing embedded monitoring system [7], such as the large size of the sensor, which destroys the original road structure after being buried. A large amount of on-site data collection is difficult to process in real time, causing communication jams and data redundancy problems. The data acquisition system is complex and heavy, with high energy consumption, which causes a lot of inconveniences when placed on site. The installation and maintenance cost of the monitoring system is too high. The monitoring data is affected by many factors, and the error of the system monitoring results may become larger as time changes [8]. However, most of the MEMS sensors for stress, strain and displacement monitoring are still in the experimental stage. There are many challenges in applying MEMS sensors to actual road monitoring. Short-term effects such as high temperature, humidity and corrosive environment in construction and long-term effects such as freeze-thaw cycle and vehicle load in actual road structure must be considered. The communication and energy supply of the sensors must also be considered.

*Corresponding Author.

To effectively solve the above problems, this paper uses the IoT technology to monitor the dynamic response information of the road under the action of the moving load of the vehicle, and realize the monitoring of the traffic and structure service status. The road acceleration sensing node collects and analyzes the road vibration response data under the moving load of the vehicle, and carries out the numerical simulation of the road dynamic response, so that can obtain the comprehensive traffic analysis of the impact of the road vibration data under the moving vehicle load on the results. Eventually, intelligent traffic control and scientific road maintenance will be realized, which can improve driving efficiency and extending road service life. It will not only significantly reduce the cost of the monitoring system, but also contribute to the popularization and application of the system.

The innovations of this paper are as follows:

1) A method for road vibration data monitoring is proposed, and a sensor node for road vibration monitoring based on the Internet of Things is developed.

2) Clustering is based on the characteristics of road vibration data, and comprehensive information monitoring methods are used to construct monitoring indicators.

3) The k-means++ algorithm is used for clustering analysis, and the vibration data indicators are domesticated, which improves the accuracy of the algorithm.

II. Related Work

A. Monitoring Technology of Road Vibration

Now, acceleration sensors have been used in the health monitoring of bridge structures. Research and practice show that vibration information can reflect the health of the structure. In the field of road engineering, the application of accelerometer is in the experimental research stage, but it shows a broad application prospect. It can be used not only for traffic information monitoring, but also for road structure state monitoring [9].

There are still many challenges in the application of acceleration sensors to road information monitoring. Due to the complex road driving environment, embedded acceleration sensors often experience data packet loss under wireless transmission. Embedded installation may cause the accelerometer battery power to be difficult to meet the long-term monitoring requirements. If the battery capacity is increased, the volume of the package node will increase and the original road structure will be destroyed. The vibration signal of the road surface under the moving load of vehicles is affected by many factors, such as axle load, vehicle speed, load acting position, node embedding depth, road structure, pavement material, vehicle suspension system, and road surface smoothness. The traditional empirical model is difficult to clarify the relationship between the vibration signal and the load on the road. Moreover, the amplitude of road vibration under vehicle load is very small. How to obtain accurate measurement values in complex interference environments, such as traffic noise, traffic in adjacent lanes, etc. How to ensure that the signal processing algorithm is simple and efficient to achieve real-time data processing and energy saving. How to package and protect the sensor so that it can withstand vehicle loads and the long-term impact of harsh environments [10]?

B. IoT Technology

With the development of time, especially with the development of computer communication technology, Internet technology, and sensor technology, the concept of the IoT has undergone considerable changes. The IoT is a kind of network that connects various networks through sensor equipment in accordance with the agreed protocol to exchange and communicate information, which can realize intelligent identification, positioning, tracking, monitoring and management [11].

It can be seen from the concept of the Internet of Things that the Internet of Things has the following characteristics: The first is a comprehensive perception, which uses RFID (Radio Frequency Identification), sensors or other sensing technologies to collect dynamic information of objects in real time. The second is to transmit the sensed information reliably in real time through various communication networks. The third is intelligent processing, using computing and other technologies to intelligently store and process massive amounts of information to realize communication between people-to-people, people-to-things, and things-to-things. These three characteristics correspond to the three-layer typical architecture of the IoT, namely, the perception layer, the network layer and the application layer. The perception layer provides perception data in the physical world, the network layer provides a link between information transmission and terminal communication, and the application layer provides diversified IoT applications [12].

Now, there are relatively few researches on the application of IoT technology to road information monitoring. The premise of the application of IoT technology to road information monitoring lies in the research and development of sensor technology in the perception layer and communication technology in the network layer, so that can obtain a large amount of perception data to serve the development and application of algorithms [13].

III. Establishment of Finite Element Model for Pavement Structure

A. Model Parameter Setting

1) Structure and material parameters: A three-dimensional finite element model of the structure and materials of the expressway is established. The pavement is mainly composed of an asphalt surface, a semi-rigid base, a sub-base and a soil base [14]. "Code for Design of Highway Asphalt Pavement" (JTGD50-2006) is referred to obtain the material parameters of each pavement structure, as shown in Table I.

The material parameters of AC-30 are estimated by referring to the lower limit value of AC-25 material parameters. Rayleigh damping is used to simulate material viscosity. The damping ratio of the pavement structure is generally between 0.02 and 0.2, and is set to 0.05 [15].

TABLE I. PAVEMENT STRUCTURE AND MATERIAL PARAMETERS

Material	Layer	Thickness of layer (cm)	Elastic Modulus Mpa	Poisson's ratio	Density kg/m^3	Damping ratio
SMA-16	Upper	4	1400	035	2400	0.05
AC-25	Middle	5	1200	0.35	2400	0.05
AC-30	Lower	7	1000	0.35	2400	0.05

2) Model size: The size of the road model (x, y, z) is 9.0m x 6.5m x 4m, the X-axis is the longitudinal direction of the road, the Y-axis is the lateral direction of the road, Z is the vertical direction of the road, and the driving direction is along the positive direction of the X-axis. The road model uses an eight-node linear reduced integral equal-parametric unit (C3D8R), and the road model is divided into 119016 units by the grid [16]. When dividing the grid, the grid in the load movement area is dense, and the distance gradually becomes sparse, as shown in Fig. 1.

The grid length*width of the load movement area is 2cm*2cm. The SMA, AC-25, and AC-30 in the surface layer are divided into two layers of grids in the vertical direction. The model is fixed around the normal direction, and the bottom is fixed in three directions. The contact between the surface layer and the base layer is completely continuous. Due to the small deformation of the soil foundation, the sub-base layer and the soil foundation are bound by Tie binding, which can appropriately reduce the number of grids and improve calculation efficiency [17]. The load movement area is the middle of the model, and the distance from the boundary is at least 3m, which reduces the influence of boundary conditions on it. The boundary conditions of the model and the load movement area are shown in Fig. 2.

B. Random Non-uniform Moving Load Setting

1) The size of the load changes with time: According to the established vehicle model, the road surface level is set as B-level road surface, and the vehicle speed is 10m/s. The random dynamic load of the vehicle on the road is obtained, as shown in Fig. 3.

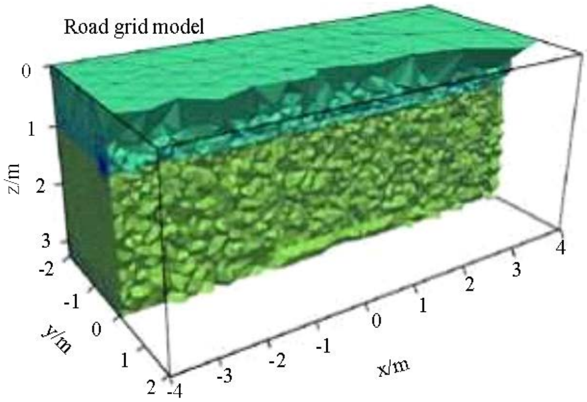


Fig. 1. Grid Road Model.

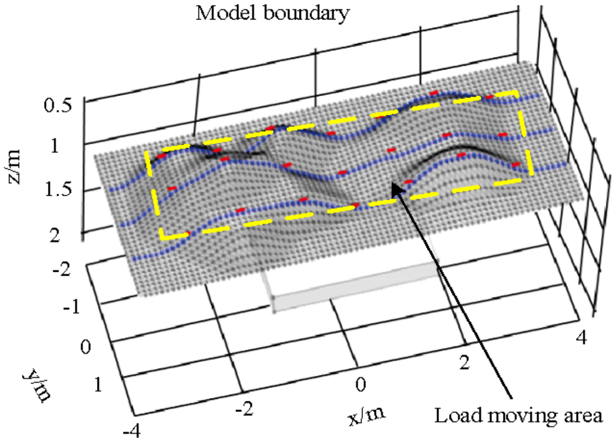


Fig. 2. Model Boundary Conditions and Load Moving Area.

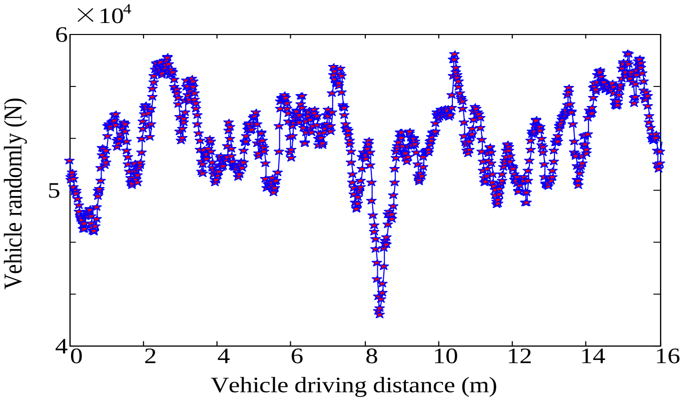


Fig. 3. Random Dynamic Load of the Vehicle.

2) The load size is distributed with space: To obtain a more realistic road dynamic response, it is necessary to consider the randomness and spatial distribution characteristics of vehicle load. After clarifying the dynamic load of the vehicle, by considering the actual contact between the tire and the road surface, a more realistic contact force distribution can be obtained. The contact between the vehicle tire and the road surface is surface contact, and its loading area can be simplified as a rectangular area [18]. In the rectangular area, the load distribution is affected by the tire tread, and the size changes with space. Fig. 4 shows the tire contacting the ground curve. The loading area of the tire on the road is a rectangular area of 20cm*18cm, which is affected by the tire pattern and forms five strip-shaped loading areas.

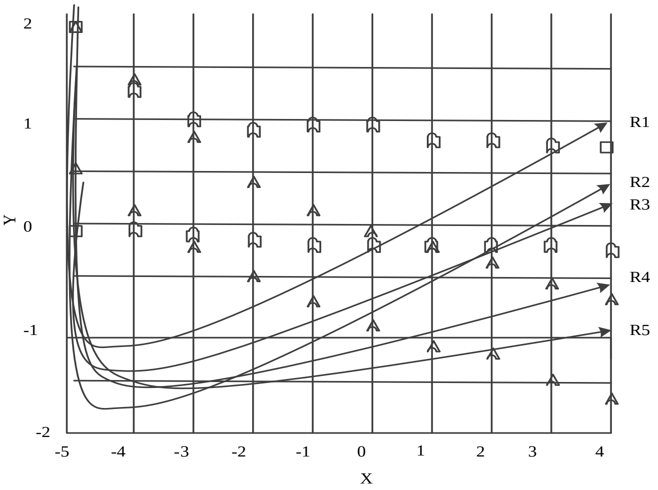


Fig. 4. The Actual Contact between the Tire and the Road.

When the vehicle moves in a straight line at a constant speed, the load amplitude ratio of the central area (R3), the side areas (R2, R4) and the edge area (R1, R5) are about 1:0.9:0.5. Moreover, in each rectangular area, the spatial distribution of the vertical load along the driving direction can be simplified to a half-sine function [19].

3) Application of Random non-uniform moving load: To simulate the moving random non-uniform load, the secondary development of the DLOAD subroutine was carried out based on the finite element software ABAQUS. Parameter analysis is as follows:

a) Sensitivity. The sensitivity of acceleration sensing node refers to the variation of sensor output voltage corresponding to the change of unit acceleration during static measurement. Therefore, the sensitivity of the acceleration sensing node can be obtained by calculating the relationship between the given acceleration and the output voltage.

b) Noise. Due to the inevitable existence of the external vibration and interference, the output signal of the sensing node includes not only the device noise of the sensing structure and the electrical noise of the signal processing module, but also a large part of the output signal is caused by external vibration and interference. Analyze the output voltage data when the sensitive axis of the acceleration sensing node is in the horizontal position.

c) Resolution. The resolution of the sensing node is the minimum variation of the acceleration that can be detected under certain conditions. The resolution can be obtained from the sensitivity and noise.

The specified coordinate function COORDS(*) and time function TIME(1) are applied to define the loading area and realize the movement of the load [20]. Equation (1) indicates that the loading area moves at a constant speed along the X-axis, and the positive direction of the X-axis is defined as the driving direction.

$$X = COORDS(1) - V \times TIME(1) - X_0 \tag{1}$$

Where COORDS(1) is the X-axis coordinate value of the integration point in the load movement area, X_0 is the initial X coordinate value corresponding to the load, V is the vehicle speed, and TIME(1) is the value of the step time [21]. Therefore, X is the X-axis coordinate value of the integration point in the load movement area corresponding to the moving coordinate system.

Equation (2) represents the spatial distribution of the load at time t, which is simplified to a half-sine function along the direction of travel.

$$Y = COORDS(2) - Y_0 \tag{2}$$

Where COORDS(2) is the y-axis coordinate value of the integration point in the load movement area, the direction of the y-axis is perpendicular to the driving direction, and Y_0 is the initial Y coordinate value corresponding to the load. $abs(X) \le \frac{b}{2}$ defines the length of the load application area, $abs(Y) \le \frac{c}{2}$ defines the width of the load application area, and α is the ratio of the load amplitude [22]. R3 is set to 1, R2 and R4 are set to 0.9, R1 and R5 are set to 0.5. b is the length of the load action area along the direction of travel, R3 is 18cm, and R1, R2, R4 and R5 are 16cm. c is the width of the load action area perpendicular to the driving direction, and the width of each small rectangle is set to 3cm. S is the actual contact area, which is the sum of the areas of R1 to R5. It is the pressure set at the integration point in the load action area corresponding to TIME(l), and is the random load value generated by the two-degree-of-freedom vehicle model [23].

The random non-uniformly distributed moving load is realized by Equation (3).

$$\begin{aligned} &If\ \ abs(X) \le \frac{b}{2} ab(Y) \le \frac{c}{2} \\ &Then\ P(t) = a \times \frac{F(t)}{S} \times \sin\frac{\pi}{b} \times X + \frac{\pi}{2}) \end{aligned} \tag{3}$$

The vehicle speed is set to 10m/s, and the length of the load movement area is set to 3m.

Therefore, the total duration is 0.3 seconds. The increment time is set to 0.001s, which is consistent with the random load sampling frequency (1000Hz). If the loading time is short enough, it can be considered that the applied load is a continuously moving load.

IV. Feature Clustering of vehicle data collection

Feature clustering is used to analyze the data of similar vehicles, which can find out the abnormal vehicle weight among similar vehicles.

1) Feature clustering: The goal of feature clustering is to find k cluster centers for the data set, so that the sum of squared distances from each point to its nearest cluster center is the smallest. Since k-means clustering randomly selects the initial cluster center, and the location of the initial cluster center will affect the clustering result. Therefore, the k-means++ algorithm is used to avoid the influence of random initialization on the result. The steps of the feature clustering algorithm are:

Step 1: A sample from the data set X is randomly selected as the initial cluster center c_I .

Step 2: A new cluster center c_I is taken, and the shortest distance is calculated between each sample $x \in X$ and the existing cluster center, that is, the distance to the nearest cluster center, denoted by $D(x)$. Calculating the probability $\frac{D(x)^2}{\sum_{x\in X} D(x)}$ that each sample is selected as the next cluster center.

Step 3: Repeating step 2 until k cluster centers are selected.

Step 4: For each sample xxx in the data set, its distance is calculated to the k cluster centers, and classified into the cluster corresponding to the cluster center with the smallest distance.

Step 5: For each category c_I , recalculating its cluster center that is the centroid $c_I = \sum_{x\in X}\frac{1}{|c_I|}$ of all samples belonging to that category.

Step 6: Repeating steps 4 and 5 until the position of the cluster center no longer changes.

2) Analysis of abnormal vehicle weight: In this paper, the cluster analysis of the two types of car models is carried out to find outliers. The analysis steps are as follows:

Step 1: Vehicle speed, temperature, and amplitude are normalized, its range is (0,1).

Step 2: K-means++ algorithm is used for clustering analysis. Because it is the same type of vehicle, the category is set to category 1.

Step 3: The coordinates of the center point are obtained, and the distance is calculated between each sample point and the center point.

Step 4: The sample points far from the center point are found out, that is, the outliers in this category.

Step 5: Since the vehicle weight is positively correlated with the amplitude, the sample point with the larger amplitude in the outlier is the point with the larger vehicle weight in the sample data.

The clustering method can be used to find outliers in the data set, which not only improves the efficiency of overweight vehicle inspection, but also reduces the energy consumption of the monitoring system. With the increase of training sample data, the accuracy of the system will be further improved, which can improve the inspection efficiency of overweight vehicles. Moreover, the camera-equipped in the system can be used as a controlling executive. When the system judges that the vehicle is abnormal, it can trigger the camera to start taking pictures instead of real-time video recording or taking pictures of all passing vehicles, which can improve the pertinence and save system energy.

3) Sensor network architecture: Using a fixed threshold for classification will increase the computational complexity and reduce the computational fault tolerance. Therefore, using the ANN (Artificial Neural Networks) for vehicle classification is easy to establish a nonlinear system model, and its programmed calculation method can reduce the complexity of the model, which is conducive to model deployment.

The ANN model has a three-layer structure, including input layer, hidden layer and output layer, as shown in Fig. 5.

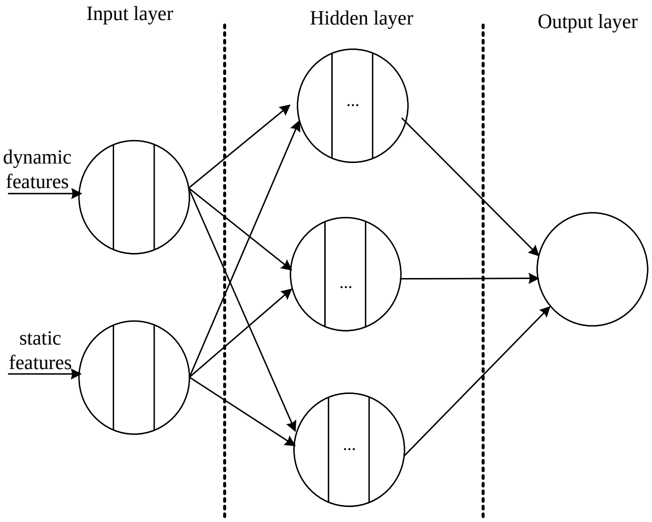


Fig. 5. Sensor Network Architecture.

The input layer contains 10 input parameters, and the vibration amplitude is the sum of the amplitude of each node in the third group. The output layer has only one node; the number of nodes in the hidden layer is set as P; to determine the optimal P value, the network with different P values is trained. When the accuracy of the trained network output is the highest, the P value is considered to be the optimal value.

V. Model Validation

To verify the rationality of the road model, the simulated data are compared with the measured data. The measured strain data comes from the fiber Bragg grating (FBG) sensor embedded on the Beijing Sixth Ring Expressway. The FBG sensor is used to measure the three-way strain response of the road surface under the moving load of the vehicle. The measured vibration data comes from the road vibration monitoring on the G320 highway. Compared to the verified monitoring points, monitoring point A is selected for comparative analysis of strain and vibration response, and monitoring point B is chosen for comparative analysis of strain response.

Wireless data transmission is also possible when the sensor platform is embedded in asphalt and concrete structures. Due to power constraints, the range of RF data transmission is limited to 40 feet. The data is transmitted through an RF wireless link located about 4 meters away from the monitoring point. The ratio of analog data to actual data is 1:100.

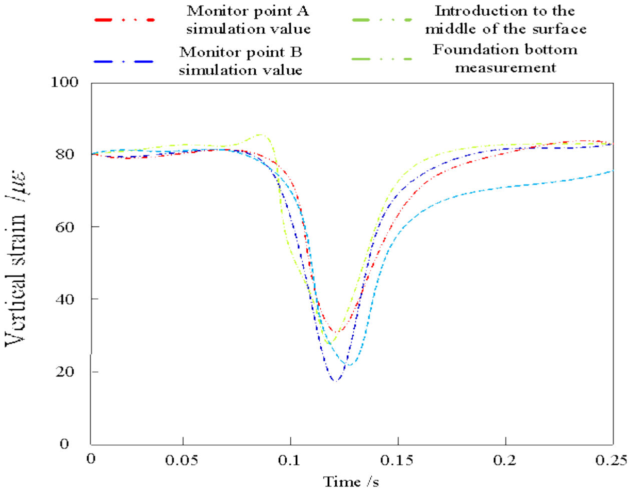


Fig. 6. Comparison of Simulated Data and Measured Data.

Fig. 6 compares the simulated data with the measured data.

It can be seen from Fig. 6 that the simulated and measured strain data match well in trend and size, and the maximum difference between the two is only about 15%. There is a deviation between the two curves, one is due to the inevitable difference between the simulated material parameters and the measured material parameters, and the other is the randomness of the vehicle dynamic load.

A. *The Stress Extreme Value Distribution in the Longitudinal Monitoring Area*

Under the situation of random non-uniform load and constant non-uniform load, the distribution of extreme longitudinal stress for pavement structure is compared. The constant non-uniform load considers the influence of tire pattern, and its contact area is the sum of the areas of R1 to R5. However, regardless of the randomness of the load, the rated load of the vehicle is adopted, and the size is 4900kN. The monitoring area is shown in Fig. 7.

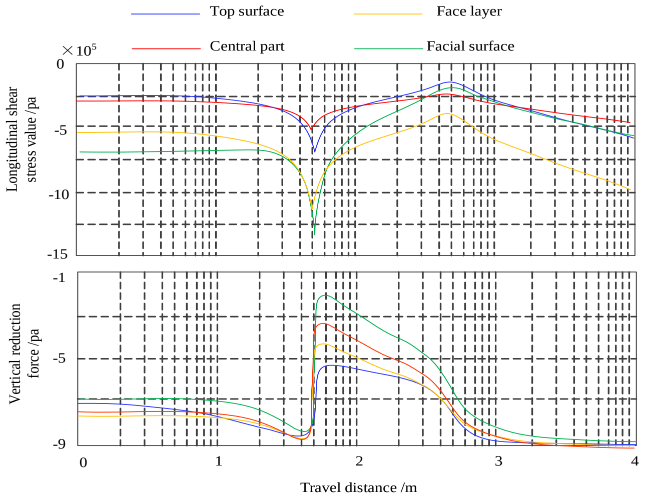


Fig. 7. The Stress Extreme Value Curve of the Longitudinal Monitoring Area of each Layer for the Pavement.

Fig. 7 shows that under the action of random non-uniform load, the extreme stress of each layer of the road surface is constantly changing, and its change characteristics are similar to the change characteristics of the vehicle's random dynamic load. However, under the action of constant non-uniform load, the extreme stress of each layer of the pavement remains unchanged. Under the action of random non-uniform load, the pavement stress fluctuation characteristics become less obvious with the increase of pavement depth. It can be seen that the random non-uniform load has a greater impact on the upper pavement structure. Moreover, under the action of random non-uniform load, the maximum stress of each layer of the pavement structure is greater than the stress extreme under the action of constant non-uniform load. And the damaging effect of random non-uniform load on the pavement is greater than that of constant non-uniform load.

B. *Distribution of Stress Extremes in the Lateral Monitoring Area*

The vertical stress extreme value is the largest when the vehicle travel distance is X=0.46m. The stress extreme value distribution of the lateral monitoring area at this position is analyzed. The horizontal monitoring area is shown in Fig. 8.

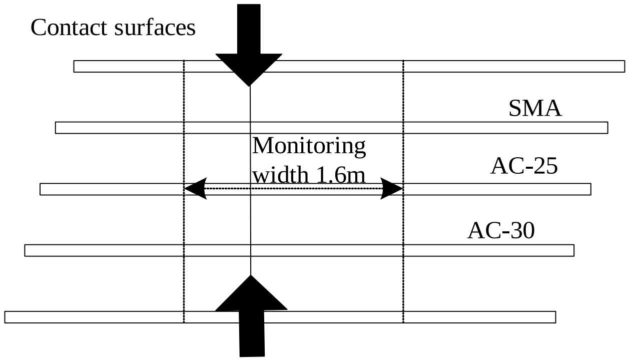


Fig. 8. Schematic Diagram of Lateral Monitoring Area at X=0.46m.

The six areas, such as the middle of SMA, the middle of AC-25, the middle of AC-30, the top of the base, the middle of the base, and the top of the base, are selected. The width of the monitoring area is 1.6m. The distribution of extreme vertical stress in the monitoring area under random non-uniform load, constant non-uniform load, and constant uniform load are compared and analyzed. For a constant uniform load, regardless of the tire pattern, the contact area is rectangular 18cm*20cm, and the rated load of the vehicle is used, and the size is 4900kN. Fig. 9 shows the distribution curve of the extreme vertical stress in the lateral monitoring area.

In comparison with Fig. 9, the uniform load has no stress concentration characteristics, while for non-uniform load, the vertical stress extreme value distribution on the road surface layer is obviously affected by the tire tread pattern. The stress distribution law is similar to the tire pattern distribution law, and the maximum stress is at the center of the tire. Secondly, there are obvious stress peaks at the stripes on both sides. As the depth increases, the overall vertical stress becomes smaller, and its spatial distribution is less affected by the tire tread. It can be seen that the spatial distribution of the load has a greater impact on the upper pavement structure.

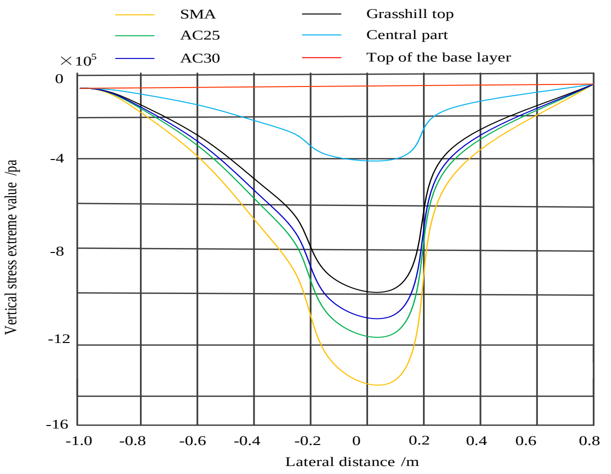


Fig. 9. The Distribution Curve of the Extreme Vertical Stress in the Lateral Monitoring Area.

Fig. 10 compares the maximum vertical stress extremes in each monitoring area.

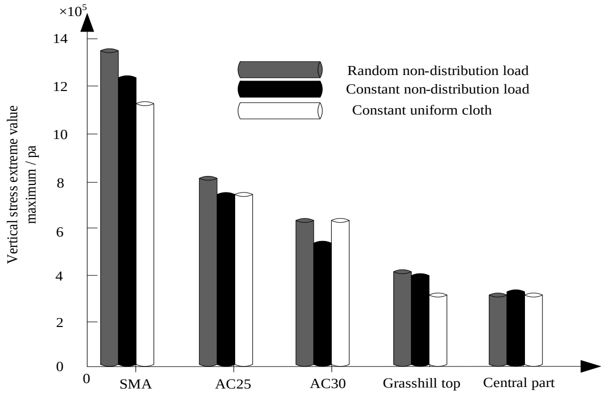


Fig. 10. The Maximum Value of the Extreme Vertical Stress in Each Lateral Monitoring Area.

Fig. 10 shows that in the middle of the SMA, the maximum stress extreme value under random non-uniform load is 32.75% higher than the maximum stress extreme value under constant uniform load. In the middle of AC25, the maximum stress extreme value in the random non-uniform load is only 11.25% higher than the maximum stress extreme value in the constant uniform load. And below the AC30 structural layer, the maximum stress extreme value of the constant uniform load is the highest, but its value is only 1.94% higher than the maximum stress extreme value under the random non-uniform load. It can be seen that as the depth increases, the difference between the maximum stress extremes for the three types of loads continues to decrease. Since the influence of the tire pattern of the vehicle, the stress distribution of the road surface layer is uneven. When the pressure is transferred to the base layer through the surface layer, the uneven stress gradually transforms into the uniform stress. In a word, if the spatial distribution characteristics of the vehicle load are ignored, the damaging effect of the vehicle load on the road surface will be underestimated.

VI. Conclusion

This paper studies the road perception nodes of road vibration monitoring based on the IoT technology, collects road vibration signals under the action of vehicle moving loads, and analyzes the characteristics and influencing factors of road vibration signals to carry out on-site monitoring. The dynamic response of the road surface under the action of random and non-uniformly distributed moving loads is simulated, and the characteristics of the road surface vibration signal under different working conditions are analyzed, and the characteristic evaluation index of the road service state is found. Finally, after the experimental analysis of the model, it is proved that different dynamic loads, surface materials and structural integrity will have a significant impact on the road vibration, and the characteristics of the road vibration signal are significantly different. In which vibration amplitude, time-domain signal waveform and frequency distribution can be used as potential evaluation indicators of road service status. Main research contents are as follows:

1) A road acceleration sensing node used for the road vibration monitoring is developed, which not only has the sensing function, but also can process, store and transmit data. The node resolution can reach 0.199 mg, can withstand the pressure of more than 67.54 Mpa, and has good waterproof packaging.

2) The road vibration amplitude is affected by the vehicle speed, vehicle weight and load position. The vehicle speed and vehicle weight are positively correlated with the road vibration amplitude, and the load position is negatively correlated with the vibration amplitude. Once the load position deviates from the monitoring point, the vibration amplitude at the monitoring point will rapidly attenuate.

3) By simulating and analyzing the dynamic response characteristics of pavement under actual loads, a method for characterizing the service state of pavement based on vibration data is proposed.

This paper realizes the road vibration monitoring under the moving load of the vehicle, and solves some technical problems of the application. In future work, the road vibration IoT monitoring prototype system can be further improved, which includes the self-powered design and performance optimization of the front-end sensing node. Its packaging and installation methods can be compatible with the characteristics of the road structure and construction technology. Through more experimental tests, the obtained monitoring results can be calibrated, and the distributed calculation and real-time processing of data can be realized.

Funding

The study was supported by “Research on roadside information perception and control technology of intelligently connected vehicles (Grant No. 2020KY25013)”.

Declaration

I declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

[1] Li M, Wang H, Xu G, et al. Finite element modeling and parametric analysis of viscoelastic and nonlinear pavement responses under dynamic fwd loading[J]. Construction & Building Materials, 2017, 141:23-35.

[2] Sarkar A. Numerical comparison of flexible pavement dynamic response under different axles[J]. International Journal of Pavement Engineering, 2016, 17(5):377-387.

[3] Zhang Y, Druta C, Wang L, et al. Dynamic responses of asphalt concrete slab under cyclic wheel loading using acceleration spectrum analysis[J]. Construction & Building Materials, 2017, 152:134-144.

[4] Huang Y, Wang L, Hou Y, et al. A prototype IOT based wireless sensor network for traffic information monitoring International Journal of Pavement Research & Technology, 2017, 11(2).

[5] Ye Z, Wang L, Xu W, et al. Monitoring traffic information with a developed acceleration sensing node[J]. Sensors, 2017, 17(12):2817.

[6] R, Coleri E, Rajagopal R, et al. Development of a cost-effective wireless vibration trucks[J]. Computer-aided Civil & Infrastructure Engineering, 2017, 32(4):433-457.

[7] Scholz T. Instrumentation for mechanistic design implementation, OTREC-RR-10-02[R]. Oregon State University, 2010.

[8] Ma W, Xing D, Mckee A, et al. A wireless accelerometer-based automatic vehicle classification prototype system[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(1):104-111.

[9] Stocker M, Ronkko M, Kolehmainen M. Situational knowledge representation for traffic observed by a pavement vibration sensor network[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(4):1441- 1450.

[10] Stocker M, Silvoncn P, Ronkko M, et al. Detection and classification of vehicles by measurement of road-pavement vibration and by means of supervised machine learning[J]. Journal of Intelligent Transportation Systems, 2016, 20(2):125-137.

[11] Jian A, Wei C5 Guo L, et al. Theoretical analysis of an optical accelerometer based on resonant optical tunneling effect[J]. Sensors, 2017, 17(2):389.

[12] Zhao H, Wu D, Zeng M, et al. A vibration-based vehicle classification system using distributed optical sensing techno logy[J]. Transportation Research Record, 2018:1-12.

[13] Zhao H, Wu D, Zeng M, et al. Support conditions assessment of concrete pavement slab using distributed optical fiber sensor[J]. Transportmetrica A: Transport Science, 2018:1-21.

[14] Dong Z, Ma X, Shao X. Airport pavement responses obtained from wireless sensing network upon digital signal processing[J]. International Journal of Pavement Engineering, 2018, 19(5):381-390.

[15] Liu W, Zhou H, Wang B, et al. A subgrade cracking monitoring sensor based on optical fiber sensing technique[J]. Structural Control & Health Monitoring, 2018(2):2213.

[16] Ma X, Dong Z J, Yu X, et al. Monitoring the structural capacity of airfield pavement with built-in sensors and modulus back-calculation algorithm[J]. Construction and Building Materials, 2017:552-561.

[17] Nabati R, Qi H. Centerfusion: Center-based radar and camera fusion for 3d object detection[C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 1527-1536.

[18] Simonelli A, Bulo S R, Porzi L, et al. Disentangling Monocular 3D Object Detection[C]. IEEE/CVF International Conference on Computer Vision, 2019:1991-1999.

[19] [19] Wang J, Lan S, Gao M, et al. InfoFocus: 3D Object Detection for Autonomous Driving with Dynamic Information Modeling[C]. European Conference on Computer Vision. Springer, Cham, 2020:405-420.

[20] Martins P F, Costelha H, Bento L C, et al. Monocular Camera Calibration for Autonomous Driving—a comparative study[C]. IEEE International Conference on Autonomous Robot Systems and Competitions, 2020:306-311.

[21] Peršić J, Marković I, Petrović I. Extrinsic 6dof calibration of a radar–lidar–camera system enhanced by radar cross section estimates evaluation[J]. Robotics and Autonomous Systems, 2019(114):217-230.

[22] Wu Y, Ogai H. Realtime Single-Shot Refinement Neural Network for 3D Obejct Detection from Lidar Point Cloud[C]. 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan, 2020:332-337.

[23] Xie Q, Lai Y K, Wu J, et al. MLCVNet: Multi-Level Context Vote Net for 3D Object Detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:10447-10456.

Automatic State Recognition of Multi-type Protection Platens based on YOLOv5 and Parallel Multi-Decision Algorithm

Ying Zhang[1], Yihui Zhang[2], Hao Wu[3], Boxiong Fu[4], Ling He[5]*
State Grid Shijiazhuang Electric Power Supply Company, Shijiazhuang, China[1, 2, 3, 4]
College of Electrical Engineering, Sichuan University, Chengdu, China[5]

***Abstract*—A protection platen is a vital component in relay protection systems. The manual inspection of protection platen states is long-term repetitive work with low efficiency and imposes a heavy burden on workers. In this work, we propose a new system to automatically detect the states of multi-type protection platens in images. This system can classify two protection platen categories and further recognize the states of protection platens. For the classification of protection platen types, we propose a new algorithm that automatically detects two protection platen types based on HSV (Hue, Saturation, Value) color space weighting operators. The proposed operators quantify the color variation in the protection platen and reduce the influence of environmental factors. With respect to the state recognition of protection platens, the Type-I protection platen states are automatically classified by the YOLOv5 (You only look once version 5) network. Since the Type-II protection platen has three primary states and more complicated structures, we investigate a new parallel multi-decision algorithm to recognize the states of Type-II protection platens based on the newly proposed watershed-color space difference-shape feature (W-CD-SF) method and the YOLOv5 network. The W-CD-SF technique can segment the protection platens and extract their shape features automatically. This multi-decision mechanism improves the robustness and generalization of state recognition. Experiments were conducted on the collected protection platen images containing 8,969 protection platens. The recognition accuracies of protection platen states exceed 95%. This system can provide auxiliary detection and long-term monitoring of protection platen states.**

Keywords—Protection platen; parallel multi-decision; YOLOv5 network; watershed-color space difference-shape feature

I. Introduction

Relay protection systems ensure the security and stability of power system operation. A protection platen is a vital component in relay protection systems [1,2]. Protection platens require manual inspection during the operation process, and such inspection is long-term repetitive work with low efficiency, which places a heavy burden on workers [3,4]. An automatic tool can help relieve the aforementioned burden on and improve inspection efficiency. With this objective, image processing techniques can be adapted to provide auxiliary detection and long-term monitoring of protection platens.

There are many types of protection platens in relay protection systems. The most commonly used types, defined as Type-I and -II protection platens are illustrated in Fig. 1. The Type-I protection platens are shown in Fig. 1(a) while the Type-II protection platens are figured in Fig. 1(b). Type-I protection platens have two categories of states, i.e., “on” and “off,” while Type-II protection platens have three categories of states, i.e., “on,” “off,” and “standby.”

In existing studies, most researchers have conducted experiments based on the images of Type-I or -II protection platens. The relevant details of previous works are the following.

1) Image processing methods for recognizing different protection platen states: Gao Jian et al. [5] eliminated the influence of shadows in images caused by illumination based on the improved Otsu threshold method and adopted the Graham algorithm to detect the states of Type-II protection platens. They achieved a detection accuracy of 96.5% for 200 shadow images. Yao Jingyan et al. [6] used the RANSAC method to restore the highlight area and evaluated the state of the Type-II protection platen by calculating the dip angle of the plate’s edges. They tested their method with 200 images and achieved an accuracy of 93%. Zhenhan et al. [7] proposed a method of automatic recognition of protection platen states based on improved sparse representation. First, they detected highlight areas in images by evaluating the maximum distances between two classes and eliminated the detected highlight areas by the improved sparse algorithm. Then, the protection platen states were recognized by their calculated minimum external rectangle. The detection accuracy of Type-II protection platens in 240 images reached 97.92%. Chen Yueqing et al. [8] proposed an improved image searching algorithm that used the background-difference and optical flow methods to extract foreground objects and recognized the Type-I protection platen switch states based on extracted edges. However, this study does not provide recognition results for specific samples. Li Tiecheng et al. [9] automatically recognized the Type-I protection platen and its corresponding text label based on an image processing technique. Combined with a model clustering and matching algorithm, the protection platens’ row number and column number in images are obtained. The recognition accuracy is obtained by comparing the calculated locations and the manually labeled locations. Fu

*Corresponding Author

This work is supported by the State Grid Hebei Electric Power Technology Project (SGTYHT/21-JS-223).

Wenlong et al. [10] recognized the Type-I protection platen states based on morphological processing. They applied their method to the images collected from different places, and most of the protection platen states could be detected. The specific detection results were not given in their research.

2) Deep-learning networks for recognizing different protection platen states: Yuan Tuolai et al. [11] used the RPN algorithm to generate a detection box and delimit the search range and then used the trained fast RCNN algorithm to detect the valuable features in this range. The accuracy of detecting the Type-I protection platen states exceeded 94%. Based on the YOLOv5 algorithm, Shi Baohua et al. [12] proposed the EYOLO algorithm, designed the local residual aggregation module, globalized the local residual characteristics, and embedded the spatial attention mechanism into the residual block to maximize the role of the local residual module. The recognition accuracy of the Type-I protection platen states reached 94.07%. Chen Xiang et al. [13] proposed an improved target detection algorithm based on a lightweight perception-v3 network. The protection platen switch states are recognized based on the migration learning strategy. The experiment was conducted on 48 Type-I protection platen images. Yang Qianwen et al. [14] proposed a bilinear fine-grained recognition method. They integrated an attention mechanism to realize end-to-end recognition of Type-II protection platen states.

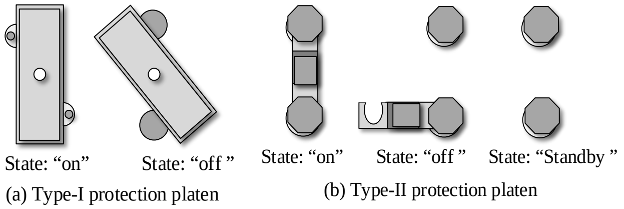


Fig. 1. Illustration of (a) Type-I and (b) Type-II Protection Platens.

The above studies illustrate that image processing methods and deep-learning networks can effectively detect protection platen states. However, these studies still have some limitations in practical applications. To overcome these limitations, we propose a new automatic state recognition system of multi-type protection platens based on the YOLOv5 and parallel multi-decision algorithms. The contributions of this work are as follows.

1) The existing studies can only recognize the states of one type of protection platen, namely a Type-I or -II protection platen. However, there are various types of protection platen in the actual operating environment. To solve this problem, we propose a new system that can classify multiple protection platen types and further recognize the states of protection platens automatically.

2) The proposed protection platen classification algorithm automatically detects two protection platen types based on HSV color space weighting operators. The proposed operators quantify the color variation in the protection platen region of interest (ROI) and reduce the influence of environmental factors through the operator weight settings of the H, S, and V color vectors.

3) A new parallel multi-decision algorithm is proposed to recognize the states of a Type-II protection platen in which parallel multi-decision is achieved by the newly proposed watershed-color space difference-shape feature (W-CD-SF) method and the YOLOv5 network. The W-CD-SF technique can segment the protection platens and extract their shape features automatically. The W-CD-SF with a support-vector-machine (SVM) classifier and YOLOv5 network classify the Type-II protection platen states in parallel, and then a fusion scheme is proposed to obtain the final detection results. This mechanism improves the robustness and generalization of the Type-II protection platen states recognition system.

II. Methods

Automatic protection platen states recognition can help relieve the burden on workers, improve inspection efficiency, and provide long-term monitoring. In this section, we detail a new system that can classify two protection platen categories and further recognize the protection platen states automatically. First, two types of protection platens are classified by the newly designed HSV color space weighting operators. Second, two states of a Type-I protection platen are automatically classified using the YOLOv5 network. Compared with the Type-I protection platen, the Type-II protection platen has three basic states and more complicated structures. In this work, a parallel multi-decision algorithm is proposed based on the proposed W-CD-SF method and the YOLOv5 network. The structure of this system is depicted in Fig. 2.

A. Automatic Classification of Protection Platen Types based on HSV Color Space Weighting Operators

Automatic protection platen states classification is an important part of power protection system inspection. There are two commonly used protection platen types, and to detect protection platen states the types of protection platens are classified first. Based on the HSV color space, an automatic classification of the protection platen category algorithm is proposed in this work, and the method is shown in Fig. 3.

1) ROI exaction of the protection platens: As shown in Fig. 3, the ROI of each protection platen is segmented in the images using the morphological method. The Sobel operator is used to calculate the image edges. Filtering and threshold processing are applied to binarize the image edge. Based on the binary image edge, the smaller connected components are removed, and the bounding box of each protection platen connected component is extracted. Through the corresponding position of the bounding boxes, the ROI of each protection platen is segmented from the original images.

2) Classification of protection platen types based on HSV color space weighting operators: To classify the types of protection platens, HSV color space weighting operators are proposed in this work. The proposed HSV color space weighting operators quantify the overall color variation degree of the ROI. The color vectors H, S, and V in the HSV color space represent hue, saturation, and value, respectively.

Environmental factors, especially illumination, have the greatest influence on value variation, second-greatest on saturation variation, and least-greatest on hue variation. In this work, the color variation degree quantization operators are proposed for the above three vectors. To reduce the influence of environmental factors on color variation quantification, the operator weights are set as follows: The operator weight of hue is set to the maximum, followed by saturation, and the weight of value is set to the minimum. The structure of each operator is shown in Fig. 4. The color variation degree quantization operators for Hue, Saturation, Value vectors are illustrated in Fig. 4(a), (b) and (c), separately.

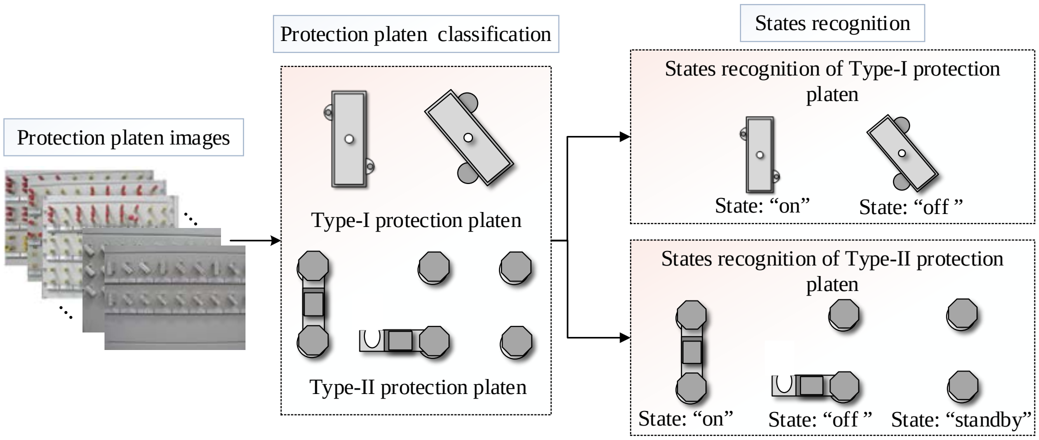

Protection platen images
Protection platen classification
Type-I protection platen
Type-II protection platen
States recognition
States recognition of Type-I protection platen
State: "on"
State: "off "
States recognition of Type-II protection platen
State: "on"
State: "off "
State: "standby"

Fig. 2. Structure of Proposed Protection Platen States Recognition System.

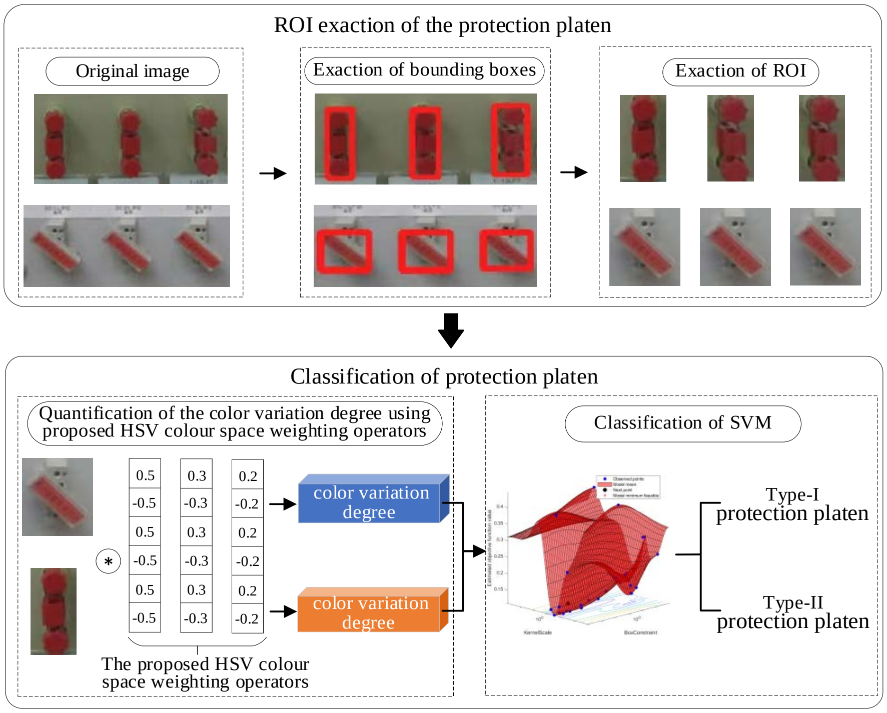

ROI exaction of the protection platen
Original image
Exaction of bounding boxes
Exaction of ROI
Classification of protection platen
Quantification of the color variation degree using proposed HSV colour space weighting operators
*
0.5 0.3 0.2
-0.5 -0.3 -0.2
0.5 0.3 0.2
-0.5 -0.3 -0.2
0.5 0.3 0.2
-0.5 -0.3 -0.2
The proposed HSV colour space weighting operators
color variation degree
color variation degree
Classification of SVM
Type-I protection platen
Type-II protection platen

Fig. 3. Diagram of Automatic Protection Platen Category Classification Method.

(a)	(b)	(c)
0.5	0.3	0.2
-0.5	-0.3	-0.2
0.5	0.3	0.2
-0.5	-0.3	-0.2
0.5	0.3	0.2
-0.5	-0.3	-0.2

Fig. 4. Structure of Each Color Variation Degree Quantization Operator: Color Variation Degree Quantization Operator for (a) H, (b) S, and (c) V Vectors.

The length of each color variation degree quantization operator is consistent with the number of ROI rows. Each ROI column is traversed using the proposed operators, and the traversal sum of all columns is the calculated ROI color variation degree. The calculation formula of the ROI color variation degree (cv_{all}) is as follows:

$$cv_{all} = \sum_{j=1}^{n}\sum_{i=1}^{m} f_H(i,j)\cdot cell_H(i) + \sum_{j=1}^{n}\sum_{i=1}^{m} f_S(i,j)\cdot cell_S(i) + \sum_{j=1}^{n}\sum_{i=1}^{m} f_V(i,j)\cdot cell_V(i) \tag{1}$$

where $cell_H$, $cell_S$, and $cell_V$ denote the color variation degree quantization operator for the H, S, and V vectors, respectively. $f_H(i,j)$, $f_S(i,j)$, and $f_V(i,j)$ represent the corresponding pixel values of the point with coordinates (i,j) in the ROI. m and n represent the index of rows and columns of the ROI, respectively.

The mean value of cv_{all} of all ROIs in the image is the calculated color variation. Based on the quantified color variation, the protective panel images are divided into two categories combined with the SVM classifier.

B. Automatic Detection of Type-I Protection Platen States based on YOLOv5

In this work, the deep-learning network YOLOv5 is used to detect the states of Type-I protection platens, which have two categories of states: “on” and “off.”

In practical applications, multiple protection platens are densely distributed on the protective panel. The target detection of the protection platens belongs to small target detection because each protective panel experimental image has many small protection platens. The YOLOv5 network uses mosaic data enhancement [15], adaptive anchor box calculation [16], and adaptive image scaling at the input layer to improve the detection ability of small targets [17]. The small target detection ability of YOLOv5 is suitable for automatic protection platen detection in this work. In addition, the lightweight structure of YOLOv5 ensures the real-time detection requirements of protection platens in practical applications.

Regarding the overall framework of YOLOv5 shown in Fig. 5, the YOLOv5 network inherits the YOLOv4 structure, including the input layer, backbone network, neck, and prediction layer. It rescales the Type-I protection platen images to 608×608 size at the input layer. The rescaled input is divided into grids, and in each grid, the coordinates of the target boxes and their confidences are predicted. The confidence of the target box can be expressed as.

$$p = \Pr(Object) * IoU_{pred}^{truth} \tag{2}$$

Each target box carries the information of (x, y, w, h, p) , where (x, y) is the relative coordinate of the normalized center point of target box, and (w, h) represents the width and height, respectively. The network outputs the corresponding prediction boxes according to the conditional probability values of the target boxes.

The input goes through the four kinds of modules of Focus, CBL, SSP, and CSPX_1 in the backbone network [18,19]. In the Focus module, the input is sliced, and then the sliced results are concatenated. The output of the Focus module is

$$Out_{Focus} = Concat(Slice(I_1, \cdots, I_n)) \tag{3}$$

where $Slice$ represents reshape operation, the input image (A, B, C) is reshaped into $(A/n, B/n, C\times n\times n)$.

The output of the CBL module is

$$Out_{CBL} = \mathrm{Re}\,LU(BN(Conv(Input_{CBL}))) \tag{4}$$

where $Conv$ represents the convolution, BN is the batch normalization, and $\mathrm{Re}\,LU$ is the activation function Leaky-ReLU.

In the CSPX_1 module, the input of the CSPX_1 module is shortcut connected:

$$Out_{CSPX\_1} = CBL(CBL(Input_{CSPX\_1})) + Input_{CSPX\_1} \tag{5}$$

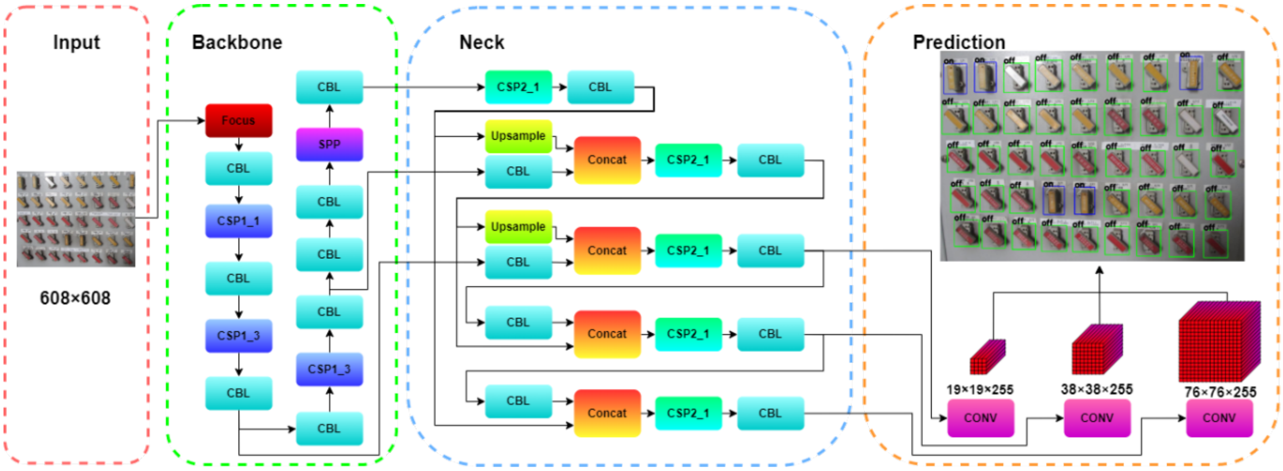


Fig. 5. Architecture of YOLOv5 for Automatic Recognition of Type-I Protection Platen States.

There are two operations in the SSP module, i.e., max-pooling and concatenation. The outputs of the SSP module are obtained as follows:

$$Out_{\mathrm{SSP}} = Concat(Maxpool(Input_{\mathrm{SSP}_1}, \cdots, Input_{\mathrm{SSP}_n})) \tag{6}$$

Through the backbone network, the feature maps of 20×20, 40×40, and 80×80 are obtained. The three sets of feature maps are combined by Neck (the structure combining FPN and PAN) for feature fusion detection. Neck outputs three feature map tensors, and the tensors are combined and passed to the prediction layer [20,21]. The prediction layer uses generalized intersection over union (GIoU) loss as the loss function and filters the target box through non-maximum suppression to obtain the final detection result. The GIoU loss is calculated as [22].

$$GIOU_{Loss} = 1 - IoU_B^A + \frac{\lfloor C(A \cup B) \rfloor}{|C|} \tag{7}$$

where A and B represent ground truth and predictions, respectively. C is the minimum closed area of A and B. Two states of Type-I protection platens are detected using the YOLOv5 network.

C. Parallel Multi-decision Algorithm for Type-II Protection Platen States Detection

The geometry and shooting background of Type-II protection platens are more complicated than those of Type-I protection platens. To achieve an accurate measurement, a parallel multi-decision algorithm is proposed based on the proposed W-CD-SF method and YOLOv5. First, the protection platens images are pre-processed to eliminate the highlight areas. Second, each protection platen is segmented based on the watershed method and color space difference information. The shape features are extracted from the segmented protection platens. Then, the states of the Type-II protection platens are automatically recognized using the proposed parallel multi-decision algorithm. The general overview of the proposed parallel multi-decision algorithm for type-II protection platen states detection is illustrated in Fig. 6.

1) Pre-processing: In this task, the quality of collected images is affected by the lighting, shooting distance, shooting angle, and other environmental factors. The highlights have a considerable influence on the protection platen segmentation. We adopt a method to detect and repair the highlight area. The details of this method are listed below.

a) Highlight area detection based on the proposed adaptive kurtosis threshold method: The detection and restoration of highlight areas are important in protection platen state detection. We propose an adaptive kurtosis threshold method to obtain a binarized mask image of the highlight areas automatically. The diagram of the proposed adaptive kurtosis threshold method is shown in Fig. 7.

Since the grey values of highlight areas differ greatly in adjacent regions of the protective panel images, the probability density curve of a grey image is adopted to evaluate the distribution of grey pixels. Generally, the highlight area is located at the back of the probability density curve due to its larger grey values. The probability density curve contains multiple peaks. The grey value corresponding to the rightmost peak is set as the adaptive threshold T_h. Based on the adaptive threshold; the binarized mask image of the highlight areas is calculated.

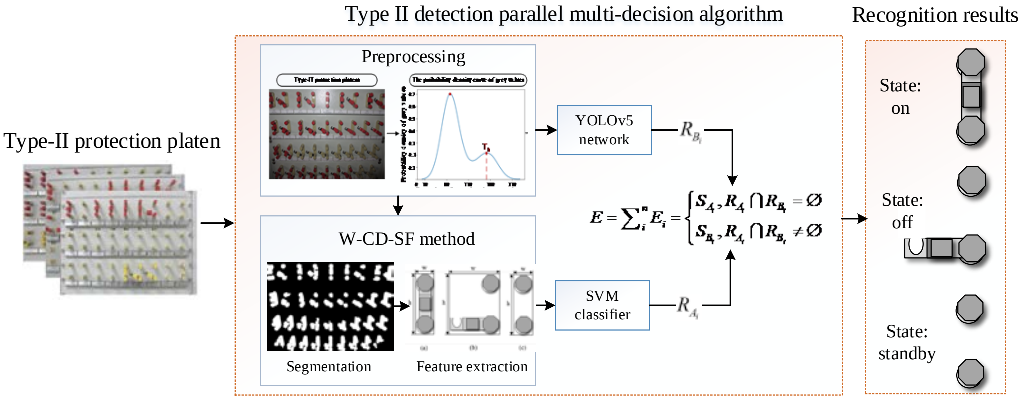

Type II detection parallel multi-decision algorithm
Recognition results
Type-II protection platen
Preprocessing
YOLOv5 network
R_{B_i}
W-CD-SF method
Segmentation
Feature extraction
SVM classifier
R_{A_i}
$E = \sum_i^n E_i = \begin{cases} S_A, R_A \cap R_B = \varnothing \\ S_B, R_A \cap R_{B_i} \neq \varnothing \end{cases}$
State: on
State: off
State: standby

Fig. 6. General Overview of the Proposed Parallel Multi-decision Algorithm for Type-II Protection Platen States Detection.

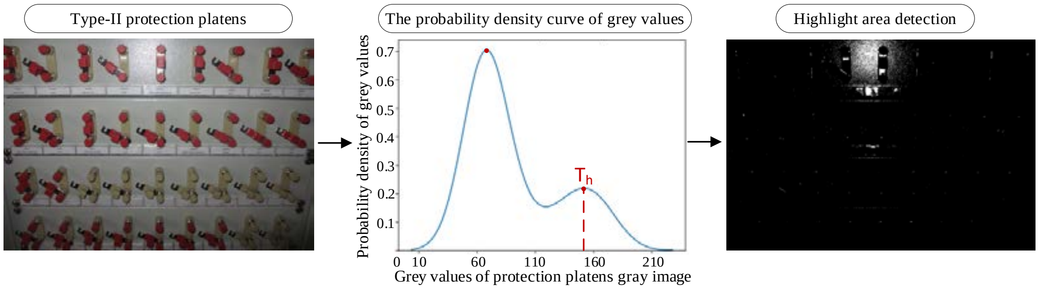

Type-II protection platens
The probability density curve of grey values
Highlight area detection
Probability density of grey values
0.7
0.6
0.5
0.4
0.3
0.2
0.1
T_h
0
10
60
110
160
210
Grey values of protection platens gray image

Fig. 7. Diagram of the Proposed Adaptive Kurtosis Threshold Method.

b) Highlight area restoration by using the fast march algorithm: The fast marching algorithm [23] is adopted to restore the detected highlight areas. The algorithm repairs the highlight areas from the boundaries of the non-zero regions in the mask image and updates the changes in the original image dynamically [24].

The fast marching algorithm fills the highlight area with pixels in its neighborhood [25]. To ensure natural filling, the weight function is adopted to distinguish the importance of different pixels in the neighborhood. The calculation formula of pixels to be filled is

$$I(p)=\frac{\sum_{q\in B_{\varepsilon}(p)}\omega(p,q)[I(q)+\nabla I(q)(p-q)]}{\sum_{q\in B_{\varepsilon}(p)}\omega(p,q)} \tag{8}$$

where $B_{\varepsilon}(p)$ denotes the neighborhood of the current pixel to be filled. $\omega(p,q)$ is the corresponding weight of neighborhood pixels, which is calculated as

$$\omega(p,q)=dir(p,q)\cdot dst(p,q)\cdot lev(p,q) \tag{9}$$

where $dir(p,q)$, $dst(p,q)$, and $lev(p,q)$ represent direction factor, geometric distance factor, and level set distance factor, respectively. The set of the above three factors ensures that the pixels closest to the normal, those closest to the pixel to be filled, and those closest to the contour of the highlight area to be repaired have the highest importance.

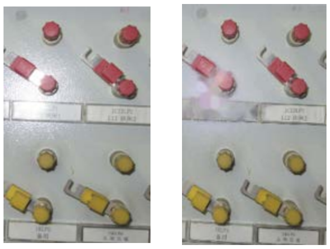

(a) (b)

Fig. 8. Illustration of Highlight Area Restoration in Protective Panel Images. (a) Original Images with Highlight Areas. (b) Restored Images.

When all the nonzero regions in the mask image are repaired, the highlight areas in the original image are also restored. An example of the highlight area restoration is shown in Fig. 8.

It is observed in Fig. 8 that the differences between the highlight area and its surrounding area are reduced.

2) Segmentation and feature extraction using the proposed W-CD-SF method: There are many protection platens in a protective panel image. To segment these protection platens, we propose a W-CD-SF method. The details of this method are introduced below.

a) Preliminary segmentation based on a watershed algorithm: The watershed algorithm is adopted to obtain the preliminary segmentation results of the protection platen. This algorithm is a transformation defined on a grayscale image. This transformation treats the image as a topographic map, with the brightness of each point representing its height, and finds the lines that run along the tops of ridges [26-29]. In this segmentation, we adopt this algorithm to obtain a continuous and close edge of each protection platen. Based on the close edge, each protection platen region can be extracted.

b) Irrelevant area elimination based on color space difference: In the second stage, we propose a new method to eliminate the irrelevant areas from segmented regions based on color difference. The irrelevant areas are mostly caused by uneven illumination and noise interference. There are color differences between the ROI and the irrelevant areas. In this method, the RGB color space is adopted to quantify the color differences. Fig. 9 shows the diagram of the proposed irrelevant area elimination method.

i) *Classification of sub-regions based on color distribution features:* Each image is divided into several 4 × 4 sub-regions. For each sub-region, the pixel value variances in the R, G, and B channels are calculated separately. The sum of the calculated pixel value variances in the three channels reflects the color distribution difference of the sub-region. Owing to the color difference between irrelevant areas and the ROI, the overall color distribution difference in the sub-region in which interference exists is greater than that in which interference does not exist. Based on the quantified color distribution difference, the sub-regions with irrelevant points are distinguished.

ii) *Interference points removal:* The pixels are arranged by their corresponding pixel value, and the first-order difference of the sorted pixel sequence difference is calculated to qualify the pixel value difference. In this work, the pixels with corresponding differences greater than the difference mean value are considered as interference points and removed.

c) Shape features extraction: To detect the states of Type-II protection platen, we extract the shape features of these segmented protection platens. These features are calculated based on the circumscribed rectangle of each protection platen's connected domain. The length h and width w of the circumscribed rectangle are obtained and illustrated in Fig. 10. Fig. 10(a), (b) and (c) correspond to the circumscribed rectangle for the state "on", "off" and "standby", respectively.

The bounding boxes of the Type-II protection platen obtained using the W-CD-SF method can be classified into two parts: protection platen region and noise. To eliminate the influence of noise and recognize the states of the Type-II protection platen more accurately, we propose the three following features with which to evaluate the shape of each bounding box:

$$HWR=\frac{h}{w} \tag{10}$$

$$CRA=h\times w \tag{11}$$

$$CRA_{mean}=\frac{\sum_{i}^{n}h_{i}\square w_{i}}{n} \tag{12}$$

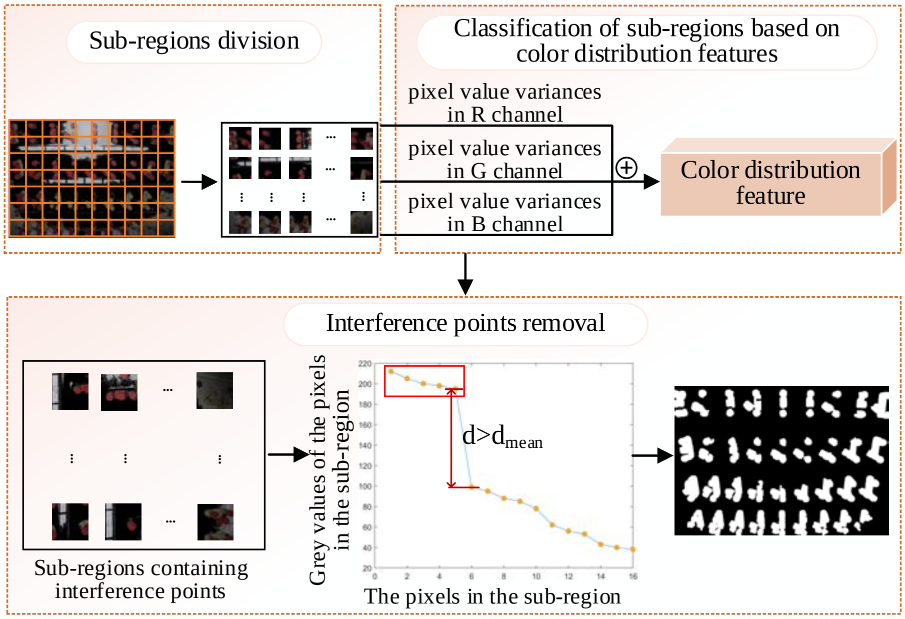

Sub-regions division
Classification of sub-regions based on color distribution features
pixel value variances in R channel
pixel value variances in G channel
pixel value variances in B channel
⊕
Color distribution feature
Interference points removal
Sub-regions containing interference points
Grey values of the pixels in the sub-region
d>d_mean
The pixels in the sub-region

Fig. 9. Diagram of Proposed Irrelevant Area Elimination Method.

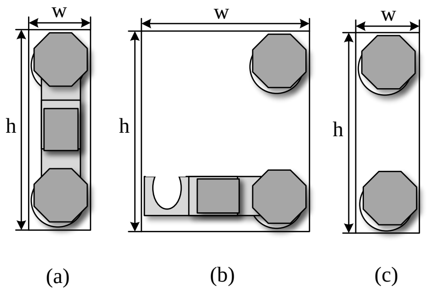


Fig. 10. Parameters Calculated from Each Connected Domain Circumscribed Rectangle: (a) State ”on,” (b) State ”off,” and (c) State ”Standby.”

where HWR is the ratio of h to w, CRA is the area of the circumscribed rectangle, and CRA_{mean} is the average area of a circumscribed rectangle of all detected objects. Finally, the Type-II protection platen states can be recognized using the segmentation masks and shape features combined with the SVM classifier.

3) Parallel multi-decision scheme based on W-CD-SF method and YOLOv5 network: The recognition result of Type-II protection platen using the W-CD-SF method is represented as A_i, which consists of the state category S_{A_i} (including states “on,” “off,” or “standby”) and the bounding box region R_{A_i}. In addition, the state recognition result of the Type-II protection platen using YOLOv5 is represented as B_i, which is also composed of the state category S_{B_i} (including states “on”, “off,” or “standby”) and the bounding box region R_{B_i}. The final state recognition result is the fusion of A_i and B_i.

Specifically, given the results of the W-CD-SF and YOLOv5 method, we can obtain the final states recognition result using the following rule:

$$E_i = \begin{cases} S_{A_i}, R_{A_i} \cap R_{B_i} = \varnothing \\ S_{B_i}, R_{A_i} \cap R_{B_i} \neq \varnothing \end{cases} \tag{13}$$

In this rule, if there is overlap between R_{A_i} and R_{B_i}, the final result is dependent on S_{B_i}, otherwise S_{A_i}.

III. Experimental Results and Analysis

The protection platen is an essential part of a relay protection inspection system. In this work, we propose a method to automatically detect the states of a protection platen. This method is composed of three parts: automatic protection platen type classification based on HSV color space weighting operators, Type-I protection platen state recognition based on YOLOv5, and Type-II protection platen state recognition based on the proposed parallel multi-decision mechanism. To evaluate the effectiveness of the proposed method, we conducted experiments on a protection platen image dataset. In this section, the dataset details, experimental settings, experimental results of automatic protection platen state recognition, along with the details of comparative experiments, are introduced and analyzed.

A. Dataset

The experimental data used in this work were collected by the State Grid Hebei Electric Power Company. There are two protection platens usually used in practical applications, i.e., Type-I and Type-II protection platens. In this work, the protective panel images containing these two types of protection platens are shot in various environments and include 3,625 Type-I protection platens and 5,344 Type-II protection platens.

B. Experimental Settings

This work achieves the automatic recognition of protection platen states. To evaluate the performance of the proposed method, the 10-fold cross-validation method is applied. The performance of the system is evaluated in terms of accuracy, false positive rate (FPR), and false negative rate (FNR). Accuracy is the proportion of correctly predicted samples to total samples. FPR is the proportion of negative samples predicted to be positive to total negative samples, which reflects the probability of negative samples being misclassified. FNR is the proportion of positive samples predicted to be negative to total positive samples, which reflects the probability of missing detection of positive samples. The calculation formulas of accuracy, FPR, and FNR are respectively, where TP denotes the number of actually positive samples predicted to be positive samples, TN represents the number of actually negative samples predicted to be negative samples, FP is the number of actually negative samples predicted to be positive samples, and FN is the number of actually positive samples predicted to be negative samples.

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+TN+FN} \tag{14}$$

$$\text{FPR}=\frac{FP}{FP+TN} \tag{15}$$

$$\text{FNR}=\frac{FN}{FN+TP} \tag{16}$$

The experiments were implemented based on the PyTorch framework. The YOLOv5 network was trained using batch size 8 for 300 epochs. The initial learning rate was set to 0.01, and the warmup technique was used for training YOLOv5. The learning rate warmup was set to 0.1. Adam was used as the optimization algorithm with a momentum of 0.8.

C. *Experimental Results of Protection Platen States Recognition*

The automatic recognition of protection platen states is a challenging task. The Type-I and -II protection platens have varied shapes. For one specific type, there are different colors of protection platens, and some protection platens and their protective panels are similar in color. Moreover, the shooting angle and distance are not the same, and the image quality is affected by lighting and shooting conditions, resulting in difficulties in recognizing the states of protection platens.

In this work, we propose utilizing the YOLOv5 network to recognize the Type-I protection platen state. Different from the Type-I protection platen, the shape of the Type-II protection platen is more complicated. To improve the recognition accuracy, in this work we propose a parallel multi-decision method combining the advantages of the proposed W-CD-SF method and YOLOv5 network.

The state classification results of Type-I and -II protection platens are listed in Tables I and II, respectively.

Table I shows that for the automatic state recognition task of Type-I protection platens the recognition accuracy, FPR, and FNR for the state "on" are 98.77%, 0%, and 1.23%, respectively. The recognition accuracy, FPR, and FNR for state "off" are 99.15%, 0%, and 0.85%, respectively. It can be seen from the results that the protection platen recognition accuracy of Type-I protection platens is over 98%, meeting the needs of practical applications.

TABLE I. TWO STATES RECOGNITION RESULTS OF TYPE-I PROTECTION PLATENS USING YOLOV5

State of protection platen	Number of protection platens	Accuracy	FPR	FNR
on	897	98.77%	0%	1.23%
off	2,728	99.15%	0%	0.85%

TABLE II. THREE STATES RECOGNITION RESULTS OF TYPE-II PROTECTION PLATENS USING PARALLEL MULTI-DECISION METHOD

State of protection platen	Number of protection platens	Accuracy	FPR	FNR
on	1,615	99.55%	8.90%	0.45%
off	2,391	96.98%	5.00%	3.02%
standby	1,338	95.33%	4.50%	4.67%

Table II lists the state detection results of Type-II protection platens using the parallel multi-decision method. It is observed that there are 5,344 Type-II protection platens, of which 1,615 are in the "on" state, 2,391 are in the "off" state, and 1,338 are in the "standby" state. As shown in Table III, for the automatic state recognition task of Type-II protection platens the recognition accuracy for the state "on" is 99.55%, that for FPR 8.90%, and that for FNR 0.45%. The recognition accuracy for state "off" is 96.98%, that for FPR 5.00%, and that for FNR 3.02%. The recognition accuracy for state "standby" is 95.33%, that for FPR 4.50%, and that for FNR 4.67%. These results show that the recognition accuracy of Type-II protection platens in three states is above 95%.

D. *Experimental Results of Type-II Protection Platen without Multi-decision Scheme*

In this work, we propose a new parallel multi-decision technique to recognize the states of Type-II protection platens. To evaluate the effectiveness of this new parallel multi-decision method, the classification results using the single-decision method with the W-CD-SF or YOLOv5 network are listed in Table III.

Table III lists the automatic state recognition results of Type-II protection platens using the single-decision method with the W-CD-SF or YOLOv5 network. As shown in Table III, for the state "on," the recognition accuracy using W-CD-SF is 93.12%, that for FPR 7.68%, and that for FNR 6.88%. The recognition accuracy using YOLOv5 is 98.88%, that for FPR 9.46%, and that for FNR 1.12%. For state "off," the recognition accuracy using W-CD-SF is 89.54%, that for FPR 10.85%, and that for FNR 10.46%. The recognition accuracy using YOLOv5 is 91.61%, that for FPR 6.19%, and that for FNR 8.39%. For state "standby," the recognition accuracy using W-CD-SF is 92.85%, that for FPR 12.08%, and that for FNR 7.15%. The recognition accuracy using YOLOv5 is 93.06%, that for FPR 11.06%, and that for FNR 6.94%. It is shown that for three states the recognition performance of each single-decision method is poorer than that of the proposed parallel multi-decision method (shown in Table II). The results indicate that the parallel multi-decision method improves the recognition accuracies by 0.68%, 5.86%, and 2.44% for “on,” “off,” and “standby” states of the protection platen, respectively.

E. Comparison Experiments

To evaluate the effectiveness of the method proposed in this work, we compared the performance of Faster-RCNN and single-shot multi-box detector (SSD) networks with YOLOv5 for Type-I and -II protection platen states recognition tasks.

Faster-RCNN is a two-stage network used for object detection tasks, especially for small objects. It has high detection accuracy and has been widely used in industry [30], [31]. The SSD network is a one-stage network that is based on regression and region proposal algorithms. It can directly predict the position and category of bounding boxes [32,33]. The results of these two deep-learning methods and our proposed method with the YOLOv5 network for recognizing protection platen states are shown in Table IV.

Table IV shows the automatic state recognition results for Type-I and -II protection platens. For Type-I protection platens, the recognition accuracy of state “on” using Faster-RCNN is 96.54% and that using SSD is 89.63%, which are both lower than the accuracy using YOLOv5 (98.77%). The recognition accuracy of state “off” using Faster-RCNN is 92.25% and that using SSD is 90.58%, which are both lower than the accuracy using YOLOv5 (99.15%). For Type-II protection platens, the recognition accuracy of state “on ” using Faster-RCNN+W-CD-SF is 97.52%, and that using SSD+W-CD-SF is 90.43%, which are both lower than the accuracy using YOLOv5+W-CD-SF (99.55%). The recognition accuracy of state “off” using Faster-RCNN+W-CD-SF is 95.46% and that using SSD+W-CD-SF is 93.65%, which are both lower than the accuracy using YOLOv5+W-CD-SF (96.98%). The recognition accuracy of state “standby” using Faster-RCNN+W-CD-SF is 96.38% and that using SSD+W-CD-SF is 89.75%, which are both lower than the accuracy using YOLOv5+W-CD-SF (95.33%). The comparison of experimental results indicates that the YOLOv5 is more suitable for automatic protection platen state classification in this work.

The experimental data were collected under various shooting environments and conditions, leading to difficulties in accurately discriminating protection platen state. The methods proposed in this work combine the advantages of YOLOv5 and image processing algorithms and perform well in the state recognition task of two protection platen types.

TABLE III. STATE RECOGNITION RESULTS OF TYPE-II PROTECTION PLATENS USING A SINGLE-DECISION METHOD WITH W-CD-SF OR YOLOV5 NETWORK

State of protection platen	Accuracy		FPR		FNR	
	W-CD-SF	YOLOv5	W-CD-SF	YOLOv5	W-CD-SF	YOLOv5
on	93.12%	98.88%	7.68%	9.46%	6.88%	1.12%
off	89.54%	91.61%	10.85%	6.19%	10.46%	8.39%
standby	92.85%	93.06%	12.08%	11.06%	7.15%	6.94%

TABLE IV. COMPARISON OF EXPERIMENTAL RESULTS

State classification accuracy of Type-I protection platen			State classification accuracy of Type-II protection platen			
Network	State of protection platen		Method	State of protection platen		
	On	off		on	off	standby
Faster-RCNN	96.54%	92.25%	Faster-RCNN +W-CD-SF	97.52%	95.46%	96.38%
SSD	89.63%	90.58%	SSD+W-CD-SF	90.43%	93.65%	89.75%
YOLOv5	98.77%	99.15%	YOLOv5 + W-CD-SF	99.55%	96.98%	95.33%

IV. CONCLUSION

An automatic protection platen state detection method was proposed in this work. In most previous research, experiments were conducted for only one type of protection platen; this was not helpful in practical applications since there are different protection platen types in the natural environment. To solve this problem, we proposed a new system that could classify multiple protection platen types and recognize each protection platen state automatically. First, the automatic classification of Type-I and -II protection platens was realized by the proposed HSV color space weighting operators. Second, the YOLOv5 network was used to detect the Type-I protection platen states. The detection accuracy for two states of Type-I protection platen was over 98%. Then, an automatic state detection system for the Type-II protection platen was presented based on the proposed parallel multi-decision method, combining the proposed W-CD-SF method and the YOLOv5 network. The recognition accuracies of the three states of Type-II protection platen (“on”, “off” and “standby”) exceeded 95%. Experimental results showed that our newly proposed system could classify multiple protection platen types and recognize each platen state automatically. In addition, the proposed system had good robustness and generalization due to the combination of a deep-learning network and image processing techniques. Future work will focus on the practical application of the proposed system and the classification and identification of other types of protection platens.

ACKNOWLEDGMENT

We thank all our participants for their contributions.

References

[1] Yuansheng, C. Qiang, X. Xiaofu, Z. Qijie, and Z. Changsheng, "An intelligent verification method for relay protection pressed board," Journal of Chongqing University, vol. 38, no. 06, pp. 91-98, 2015-12-15. 2015.

[2] Y. Yu, C. Fu, C. Baiqing, Z. Meiyong, Z. Yiran, and Y. Chao, "Research on substation protection hard plate detection and state recognition technology based on deep learning," Technology Innovation and Application, vol. 11, no. 24, pp. 25-29, 2021-08-28. 2021.

[3] X. He and Y. Wang, "Intelligent Recognition System of Substation Hard Platen State Based on Machine Learning," International Conference on Power System Technology (POWERCON), IEEE, 2018, pp. 4320-4325.

[4] R. Wu, W. Zhang, H. Chen, and J. Jiao, "Research on State Recognition of Platen Based on Improved K-means Algorithm," 2020 International Conference on Electrical Engineering and Control Technologies (CEECT), IEEE, 2020, pp. 1-7.

[5] G. Jian, Y. Shiyong, S. Zhengyu, Y. Zheng, L. Zhenhan, and Y. Jingyan, "Research on the identification of protection press plate state based on OTSU-Graham improved algorithm," Electrical Measurement and Instrumentation, pp. 1-9, 2021-05-19. 2021.

[6] Y. Jingyan, S. Zhengyu, G. Jian, and L. Zhenhan, "Research on operation state identification method of protection platen based on image fusion," Electrical Measurement & Instrumentation, vol. 58, no. 08, pp. 88-96, 2021-03-08. 2021.

[7] L. Zhenhan, S. Zhengyu, G. Jian, and Y. Jingyan, "Status identification of substation protection plate based on improved sparse representation," ELECTRONIC MEASUREMENT TECHNOLOGY, vol. 44, no. 23, pp. 86-92, 2021-12-30. 2021.

[8] C. Yueqing et al., "Condition Recognition System for Relay Protection Plate Based on Improved Bag of Feature Algorithm," Power Grid Analysis & Study, vol. 49, no. 02, pp. 99-106, 2021-02-20. 2021.

[9] L. Tiecheng, R. Jiangbo, L. Qingquan, Z. Yuhao, W. Zhihua, and H. Yanjiao, "Image Recognition and Model Cluster Matching of Relaying Plate," JOURNAL OF HARBIN UNIVERSITY OF SCIENCE AND TECHNOLOGY, vol. 26, no. 04, pp. 70-77, 2021-08-27. 2021.

[10] F. Wenlong et al., "Protection platen status recognition based on image processing and morphological feature analysis for smart substation," Electric Power Automation Equipment, vol. 39, no. 07, pp. 203-207, 2019-07-12. 2019.

[11] Y. Tuolai, L. Xinhai, L. Haixin, Z. Lingcheng, M. Chenxu, and Y. Yanhe, "State identification method of relay protection pressing plate based on fast r-cnn algorithm," Electrical Technology and Economy, no. 06, pp. 36-39, 2021-12-20. 2021.

[12] S. Baohua, J. Renyue, Y. Chao, L. Zhenxing, and L. Yanzhang, "Enhanced YOLO network for status recognition of a substation protection plate," Power System Protetion and Control, vol. 49, no. 23, pp. 163-170, 2021-12-01. 2021.

[13] C. Xiang, Z. Qingnian, X. Shaoyu, and C. Cuiqiong, "Identification of Platen Switch State Based on Transfer Learning Strategy," JISUAN]I YU XIANDA IHUA, no. 05, pp. 120-126, 2021-05-15. 2021.

[14] Y. Qianwen and Z. Ke, "Press-Plate State Recognition Based on Improved Bilinear Fine Grained Model," Laser & Optoelectronics Progress, vol. 58, no. 20, pp. 146-155, 2021-10-25. 2021.

[15] D. Wang and D. He, "Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning," Biosyst. Eng., vol. 210, pp. 271-281. 2021.

[16] S. W. Kang and U. S. Choi, "ROI Image Encryption using YOLO and Chaotic Systems," INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, vol. 12, no. 7, pp. 466-474, 2021-01-01. 2021.

[17] C. Dewi, R. Chen, and H. Yu, "Weight analysis for various prohibitory sign detection and recognition using deep learning," Multimed. Tools Appl., vol. 79, no. 43-44, pp. 32897-32915. 2020.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," IEEE T. Pattern Anal., vol. 37, no. 9, pp. 1904-1916. 2015.

[19] J. Liu and D. Zhang, "Research on Vehicle Object Detection Algorithm Based on Improved YOLOv3 Algorithm," Journal of physics. Conference series, vol. 1575, no. 1, p. 12150, 2020-01-01. 2020.

[20] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117-2125. 2017.

[21] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8759-8768. 2018.

[22] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 658-666, 2019-01-01. 2019.

[23] X. L. Huan, H. Zhou, and J. L. Zhong, "LSB based Image Steganography by using the Fast Marching Method," INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, vol. 10, no. 3, pp. 1-5, 2019-01-01. 2019.

[24] S. Wang, X. Yan, F. Ma, P. Wu, and Y. Liu, "A novel path following approach for autonomous ships based on fast marching method and deep reinforcement learning," Ocean Eng., vol. 257, p. 111495. 2022.

[25] Y. Qi, "Multi-stencils fast marching method for factored eikonal equations with quadratic anisotropy," Appl. Math. Comput., vol. 417, p. 126776. 2022.

[26] V. Sivakumar and N. Janakiraman, "A novel method for segmenting brain tumor using modified watershed algorithm in MRI image with FPGA," Biosystems, vol. 198, p. 104226. 2020.

[27] A. M. Anter and A. E. Hassenian, "CT liver tumor segmentation hybrid approach using neutrosophic sets, fast fuzzy c-means and adaptive watershed algorithm," Artif. Intell. Med., vol. 97, pp. 105-117. 2019.

[28] L. Zhang, L. Zou, C. Wu, J. Jia, and J. Chen, "Method of famous tea sprout identification and segmentation based on improved watershed algorithm," Comput. Electron. Agr., vol. 184, p. 106108. 2021.

[29] W. Cao, Z. Qiao, Z. Gao, S. Lu, and F. Tian, "Use of unmanned aerial vehicle imagery and a hybrid algorithm combining a watershed algorithm and adaptive threshold segmentation to extract wheat lodging," Physics and Chemistry of the Earth, Parts A/B/C, vol. 123, p. 103016. 2021.

[30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans Pattern Anal Mach Intell, vol. 39, no. 6, pp. 1137-1149, 2017-06-01. 2017.

[31] R. F. Mansour, J. Escorcia-Gutierrez, M. Gamarra, J. A. Villanueva, and N. Leal, "Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model," Image Vision Comput., vol. 112, p. 104229. 2021.

[32] J. Kim, M. Jung, and J. Kim, "Decoupled SSD: Reducing Data Movement on NAND-Based Flash SSD," IEEE Comput. Archit. L., vol. 20, no. 2, pp. 150-153. 2021.

[33] Z. Shen, L. Han, C. Ma, Z. Jia, T. Li, and Z. Shao, "Leveraging the Interplay of RAID and SSD for Lifetime Optimization of Flash-Based SSD RAID," IEEE T. Comput. Aid. D., vol. 40, no. 7, pp. 1395-1408. 2021.

Dual Authentication for Payment Request Verification Over Cloud using Bilinear Dual Authentication Payments Transaction Protocol

A. Saranya[1]
Research Scholar, Department of Computer Science and Engineering
SRM Institute of Science and Technology
Kattankulathur, Chengalpattu, Chennai, Tamil Nadu, India

R. Naresh[2]*
Associate Professor, Department of Networking and Communications
SRM Institute of Science and Technology
Kattankulathur, Chengalpattu, Chennai, Tamil Nadu, India

***Abstract*—There has been a recent explosion in the number of mobile network payment gateways that enable users to access services through a variety of devices. Mobile payment gateway security is complicated by a number of difficult-to-solve issues. As digital technology has progressed over the last decade, mobile payment mechanisms have gained a lot of interest. In the internet industry, these standards might have a significant impact on service quality. However, the most important aspect to consider when using these systems is their accountability, which assures confidence between the parties engaged in the financial transactions. Mobile payments may be easy, quick and secure. On the other hand, they may be rather pricey and are still susceptible to problems caused by technology. Specifically, mobile payments won't be able to go through at all if there are any problems with the host phone. For this reason, in this article a mobile payment mechanism that uses secure bilinear dual authentication. Using Bullet hash Maximum distance separable (MDS) and the mutate Hellman algorithm, our payment protocol incorporates all of the essential security characteristics to establish confidence between the parties. To put it another way, accountability is assured by mutual authentication and non-repudiation. Conflicts that may emerge in the course of a payment transaction may be resolved using our strategy. Scyther is used to test our suggested protocol's empirical performance.**

Keywords*—*Mobile payment; transaction protocol; bullet hash maximum distance separable; mutate Hellman algorithm

I. Introduction

Mobile smart devices (such as smartphones or laptops) have become more common in everyday life as a result of the development and widespread usage of mobile communication technology [1]. This results in an ever-increasing amount of online service requests. Online services such as Ali Pay, Apple Pay, We Chat Pay and so on rely on mobile payments, which are attracting a lot of attention. These online transaction programs let customers to purchase a wide range of goods and services from wherever they have internet access. A user's private identifying information is often made available to retailers when an online transaction is initiated in order to verify the transaction's authenticity. The unreliability and avarice of merchants may lead them to offer items that the user does not need or merely to sell the identity of the user for economic gain to other parties. There must be an ability to validate a transaction message's legality and validity, so that the merchant may ensure that products or services are given by the proper user. [2] Verification of transaction messages may also prevent users from claiming that they didn't acquire the products and services they claimed to have purchased. Many cryptographic primitive-based protocols for mobile payments have been developed to meet these security requirements. Their protocols meet security standards such as user anonymity and unforgivability. The identity of any merchant or opponent cannot be linked to a transaction message when a protocol maintains anonymity for the users involved [3]. Unforgeability, on the other hand, implies that the sender of a message can be identified and that anybody attempting to forge the transaction message of another party will be caught [4]. A mobile payment system must take efficiency into consideration in addition to security considerations. As a result, it is important that a payment protocol only need a minimal amount of processing power and storage space to be implemented on low-powered gadgets. Traditional transaction protocols provide user public key certificates through a public key infrastructure (PKI) [5, 6]. Verifying the authenticity of a public key may be done with the help of a trusted certificate authority. Because of the need for extra resources to manage certificate revocations, distribution, and storage, the use of PKI significantly raises the costs of both communicating and storing data. As a result, there is a contradiction between PKI and the capabilities of mobile devices with limited processing and storage. It is still difficult to build a mobile transaction protocol with minimal resource requirements in terms of computation, network traffic, and data storage. Addressing the aforementioned issue, a novel mobile payment system was created that simultaneously guarantees secrecy, immutability, and minimal resource usage. The following is a brief summary of its most important contributions:

1) First, the dual authentication payment transaction method was provided.

2) A mobile payment mechanism based on our suggested approach is demonstrated. Furthermore, Pay Platform serves as a trusted proxy for users when they want to securely communicate with Merchant Server. As a result, users can rest easy knowing that they won't be sending or receiving private messages from merchants. Since most calculation takes place on Pay Platform, user resource utilization is lowered as well. Unforgeability is ensured by the use of certificate less public key cryptography and signatures on every piece of transaction data.

3) It is important to maintain the Pay Platform and Merchant Server as lightweight as possible, despite the fact that they must do computations for each transaction, in order to solve the scalability problem. When processing a payment on the Pay Platform or the Merchant Server, the signature verification process is by far the most time-consuming step.

4) The protocol was implemented to test the using the Bullet hash MDS, a mutated version of the Hellman algorithm, and various mobile payment protocols. Comparison reveals that our protocol is viable and efficient in terms of secure payments.

Rest of the paper has been divided into sections. Section II focuses on other studies in this field. The problem statement is presented in Section III. Section IV demonstrates the suggested technique's terminology in action. Section V focuses on experiments and implementation. Towards the end of the article, the conclusion is reached.

II. Related Work

Mobile payment systems have been the focus of a number of researches in recent years aimed at strengthening their security. Many of these efforts have been directed to developing foundations for the establishment of new electronic payment systems, as well. Username and passwords, symmetric and non-symmetric and elliptic curve encryption, smart cards, 2D bar codes, and biometric technologies have all been tried for electronic payment system authentication. All sorts of authentication techniques and protocols are based on these notions. Symmetric and asymmetric signatures fail in M-commerce despite being frequently utilized for authentication. In [7], mobile payment solutions based on durable certificate-less signatures and bilinear pairing are proposed. [7] To make the suggested mobile payment system acceptable for mobile devices with low processing capacity, they smartly improve it. Security and performance testing of the suggested mobile payment system on the Raspberry PI have demonstrated its feasibility. Anonymous and untraceable payments have never previously been feasible on a mobile device according to a novel payment technique explained in [8]. Because mobile devices have limited computing capacity, the proposed protocol depends on a Pay Platform to handle the majority of the computational burden (which is almost usually equipped with lots of processing power) (which is almost always equipped with plenty of processing power). Batch-verification has been implemented to lessen the overhead for millions of users on the Pay Platform and Merchant Server because the Pay Platform and Merchant Server must run calculations for each transaction. In terms of cloud security and privacy, one of the most pressing challenges is whether or not sensitive information may be accessed by other parties. The researchers [9] have presented cloud service providers with an intelligent encryption solution that secures cloud service providers' access to the incomplete data. Dispersed cloud servers are recommended for data storage after the data file has been partitioned. SA-EDS is a notion and an algorithm that underpins the Secure Efficient Data Distribution (SED2) Method, the EDCon Methodology, and the Alternative Data Distribution (AD2) Algorithm. ZeroMT is a method presented in [10] that enables several simultaneous covert balance transfers. This method ensures that the balances in all accounts and the amounts transferred remain private." "Since all transfers are contained within a single transaction, there are fewer transactions to verify on the main chain. Off-chain, they are the standard building block for a multi-transfer transaction that is undetectable by unauthorised parties but can be validated by smart contract infrastructures. NFC devices and payment terminals connect with each other using this protocol, which is based on the EMV standard [11]. EMV's weaknesses will be solved by adding an extra layer of security and beefing up the core EMV communications, according to the protocol. Due to the lack of a reliable third party, security features such as the detection of double spending and the prevention of brute force attacks are still in development. In order to make secure mobile payments, NFC-enabled cellphones may be utilised, as stated in [12]. Abughazalah's protocol has the potential to address issues of privacy and security. In order to ensure the security of this protocol, a one-time password (OTP) system was implemented. The OTP mechanism is preinstalled in NFC-enabled smartphones. However, without a trusted third party, this protocol lacks essential security features like the ability to prevent double spending and ensure all participants are treated fairly. According to [13], mobile vouchers for NFC-enabled phones are safe to use. The suggested strategy focuses largely on the collection and redemption of loyalty points. However, there are still several potential security holes in this system, including those related to message privacy, mutual authentication, and detecting duplicate spending. This results in an inherently unfair implementation of the protocol. An EMV-compliant near-field communication (NFC) mobile payment system was presented by the author in [14]. Data and digital signatures are secured by cryptographic methods like symmetric key encryption and public key encryption in their proposed protocol. Unfortunately, this protocol is vulnerable to a wide range of attacks despite the fact that it is not fair and cannot rely on a neutral third party to ensure its security. In [15], the author presented a secure contactless NFC payment system using NFC bank cards and an internet connection with a respectable organization. Client payment devices limited to NFC bank cards that do not need Wi-Fi or 4G. Due to the absence of a reliable third party, this protocol is not yet capable of detecting duplicate spending or protecting against brute-force attacks, and it is therefore not yet fair. An NFC mobile payment system with cloud-based security was shown in [16]. The usage of NFC radio waves in a public setting makes it more secure than the EMV standard. Other security flaws, such the ability to spot double spending, may be exploited even in the absence of a third party that can be trusted. The au's anonymous mobile payment system, which was introduced in [17], has increased the safety of mobile commerce data. They assert that their protocol satisfies all of the necessary conditions for non-repudiation, anonymity, and unlinkability. However, there is no safeguarding against message corruption or brute-force attacks in this protocol [18-21]. The trustworthy third party's knowledge of sensitive transaction data like payments and invoices also poses a security risk. Using key distribution, the author of this research presents [22] a cloud-based efficient authentication system for mobile payments. Our innovative certificate-less proxy re-signing approach for mobile payments not only protects your privacy, but also simplifies your data storage needs. Using the Secure Authentication Protocol, [23] suggests a mobile payment system (SAP). By using cryptographic methods, it was able to create a reliable solution that can identify fake servers and clients. While the use of mobile devices in the payment industry continues to rise, the proposed method ensures the safety of user accounts and individual privacy. Smart mobile device makers and mobile data users both have a role to play in resolving this problem so that mobile payments may continue to be used in public communication networks that are not always secure. They provide a secure business strategy for mobile payments in the e-healthcare application by making use of key distribution cryptographic techniques. In [24], a novel, effective trust model for secure, dynamic group communication across distant networks is presented. The concept of a hierarchical attribute-based encryption was introduced by the author in [25]. A single encryption key may secure a whole directory's worth of data. The computational and storage requirements of their attribute-based hierarchical file encryption method are less than those of prior research. A novel index model, the data vector (DV) tree, based on a crossover genetic process, was developed with the help of soft computing. The file's term and inverse document frequencies, together with any other features included in the file, are used to build the DV tree. Their new key generation, encryption, and decryption tool is an extra bonus for anybody using their data. The authors of [26] proposed a machine learning algorithm to identify malicious uniform resource locators by combining URL lexical selections, payload size, and python supply parameters. A Chaotic Hopfield Neural Network combined with an adaptive encoding approach may provide a more secure model for keeping private information; the proposed approach enhances the safety of a shared key among any number of nodes [27-32].

A. Research Gap

Consequently, the existing solutions for mobile payment protocols lacked both substantial data security and fairness. There is no credible third party under the constraints of such procedures. To avoid wasting time and money in the event of a disagreement, no one gathers and records all transactional data. This means that the processes at play cannot ensure a fair outcome. As a result, there is potential for a variety of attacks, including physical force, man-in-the-middle, and double spending. In this work, a strategy for resolving these issues is demonstrated. That is to say, the proposed protocol simultaneously guarantees both high levels of fairness and critical data security. The offline session key generation and hashing features are used for further security and mobility. Basically, it was impossible to stop repetition or brute-force attacks by never reusing the session keys that are created for each and every transaction.

III. Problem Statement

No one will trust m-commerce until there is a secure method of exchanging business information and conducting electronic financial transactions over mobile networks. Secure connections and transfers are essential for M-payment systems, which send sensitive data. As a result, high levels of perceived and technological security are required for m-payment systems to be widely used and widely accepted by customers. M-commerce has seen the implementation of a number of different mobile security and payment mechanisms. Cryptography techniques are essential to satisfy the above-described transaction security standards. The security of mobile payments done via networks with little or no physical protection is greatly enhanced by these devices.

IV. Proposed Work

Client, Merchant, Acquirer, Issuer, and Payment Gateway are all parts of the BDAPTP architecture was developed (which is in turn based on the Client Centric model) (all of which were described earlier). Below are three examples of simple Payment transactions that illustrate the Client Centric approach in action.

- When a customer completes a purchase, the money is transferred to the shop owner.
- With Value Subtraction, the Client's funds are deducted by the Payment Gateway (on behalf of the Issuer).
- The Acquirer's Payment Gateway sends funds to the Merchant's Value Claim account. When using a Payment Gateway, you may get in touch with an Acquirer straight away (an entity that provides infrastructure for a Merchant to take credit card and other kinds of electronic payment). The suggested method has no need for communication between the merchant and the acquirer. A wireless or cellular network provided by a mobile phone provider is used to create an Internet connection between the Client and the Payment Gateway. The overall illustration of the suggested methodology was depicted in the Fig. 1.

A. BDAPT Protocol

There are a variety of processes for payment transaction depicted in Fig. 2. An individual identification number was assigned to each mobile phone. If an IMEI number is placed on a blacklist by a wireless network operator, GSM, it is used to identify legitimate devices on the network and to prevent ongoing usage of a phone that has been blacklisted from using the network. Different aspects of the suggested mobile payment security were identified in this article. The stages involved are:

1) Registration phase: Users and merchants must register on the payment gateway in order to get legitimate credentials, such as a mobile PIN for login and the encryption key, which must be generated.

2) Transaction phase: To make a payment, a user must first send a payment request to the merchant, which the merchant must confirm by verifying the user's payment information and personal information.

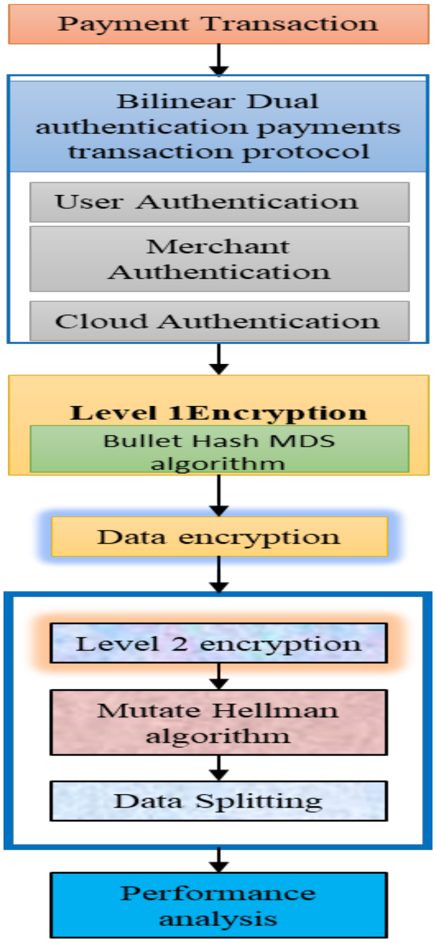


Fig. 1. Schematic Representation of the Suggested Methodology.

3) *Authentication phase:* The Issuer and Acquirer calculate the trust value to authenticate the payment credentials and approve the payment via the confirmation from the merchant module, which is then executed on the payment gateway.

It shows the process of the merchant server authentication using the proposed protocol. Merchant and admin servers must agree to the process of generating and disseminating a secret key for data encryption and decryption during authentication in order for the suggested method to operate. The Merchant Server employs "bank number" and a secret key to authenticate the admin server, creating a basic ciphertext when the authentication process is complete. The cipher text formula for merchant server authentication is:

$$\text{Merchant Server Authentication (MSA)} = \text{bankno (xor) trust value} \quad (1)$$

In order to avoid a session hijacking on the buyer's end, each authenticating demand should be issued a distinct private key.

$$\text{Admin Server Authentication (ASA)} = \text{MSA (xor) trust value} \quad (2)$$

$$\text{Admin server trust value (ASV)} = \left(g^{bankno}\right) \quad (3)$$

$$\text{Merchant share value(MSV)} = \left(g^{H(bankno_{no})}.g\right)^{bankno} \quad (4)$$

Whenever a client sends a payment, ask for the customer's transaction number in the cloud. Processing the request data requires the use of a trapdoor-creating method for the simple reason that if the payment requests were transmitted over the plain text, we run the risk of them being tracked and attacked. As a result, enhancing security and preventing the attack may be accomplished via the trapdoor design process.

$$\text{Patient Number Request (PNR)} = \left(g^{H(bankno_{no}).pn}, g^{pn}\right) \quad (5)$$

where $PN1 = g^{H(bankno_{no}).pn}$ and $PN2 = g^{pn}$

The cloud should now check and validate the merchant data and the mobile user request, as shown below, after it has received the whole request, trusted cloud exchanged between merchant server and mobile user across a distribution channel.

$$\text{Cloud payment request matching} = e\ (PNR, MSV) \quad (6)$$

$$= e\left(\left(g^{H(bankno_{no})}.g\right)^{bankno}, g^{pn}\right) \quad (7)$$

$$= e(g,g)^{(H(bankno_{no})+1).bankno.pn} \quad (8)$$

$$= e(g,g)^{H(bankno_{no}).bankno.pn}.e(g,g)^{bankno.pn} \quad (9)$$

$$= e\left(g^{H(bankno_{no}).pn}, g^{bankno}\right).e\left(g^{bankno}, g^{pn}\right) \quad (10)$$

$$e\ (PNR, MSV) = e(PN1, ASV).e(ASV, PN2) \quad (11)$$

A user's User Id is required before they may initiate a cloud transaction. The system will direct the user to the registration page if they have not previously done so. Once a user's id is found in the bank's database, the next step is to have them input a password. It is verified and if found to be inaccurate, the user is led back through the process, or if it is correct the user will be sent to a new step in which the system will ask for a unique identification (UID). For 24 hours, the user's account will be suspended if the incorrect UID is entered. If the user's UID and QR Code are matched, an OTP will be sent to the user's registered cellphone number; if the user's UID and QR Code are not matched, the user's account will be blocked for 24 hours. Once the OTP has been entered, the following step is to verify that the OTP entered is accurate. If the OTP is accurate, the key is kept secret throughout the encryption step. For 24 hours, the account will be blocked if the OTP entered is wrong. If the OTP is entered correctly three times, the system will send you to the right OTP entry procedure. All input is encrypted using the Bullet hash MD5 algorithm, a public key cryptosystem that provides a hash key. To further increase the degree of security in the suggested technique, the asymmetric public key cryptosystem modify Hellman algorithm was used to generate the hash key. Data is subsequently divided into 128-bit XOR operations and sent to the server for further processing.

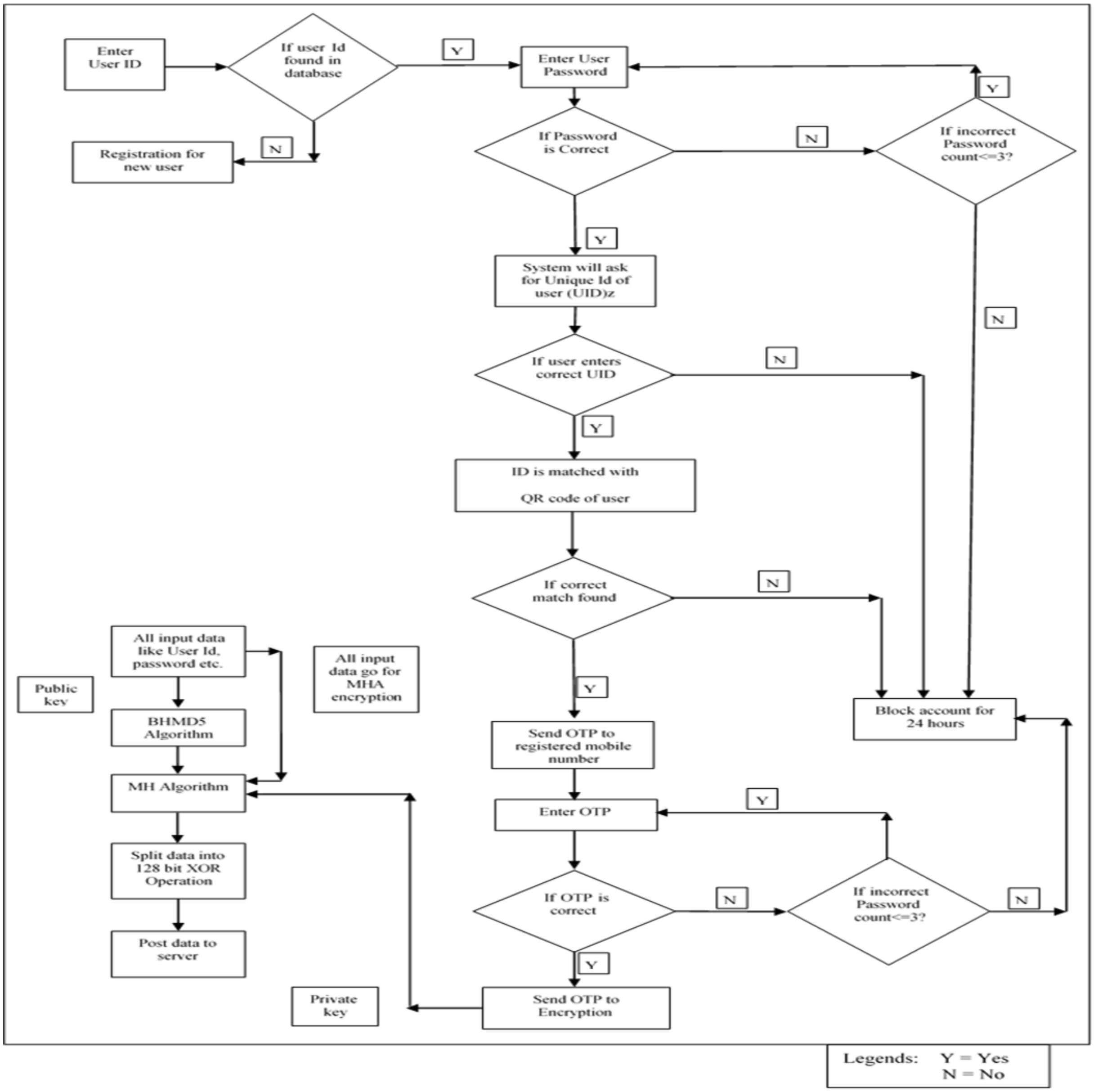

Enter User ID
If user Id found in database
Y
N
Registration for new user
Enter User Password
If Password is Correct
N
If incorrect Password count<=3?
Y
N
Y
System will ask for Unique Id of user (UID)z
If user enters correct UID
N
Y
ID is matched with QR code of user
If correct match found
N
Y
Block account for 24 hours
All input data like User Id, password etc.
All input data go for MHA encryption
Public key
BHMD5 Algorithm
MH Algorithm
Split data into 128 bit XOR Operation
Post data to server
Send OTP to registered mobile number
Enter OTP
Y
If OTP is correct
N
If incorrect Password count<=3?
N
Y
Private key
Send OTP to Encryption
Legends: Y = Yes
N = No

Fig. 2. Process of Transaction.

B. Units Cryptographic Transaction

1) Bullet hash MD5 algorithm: Integrity security is mostly provided via hash functions. Authentication is also provided when they are used in conjunction with digital signature and message authentication code (MAC) techniques. As an example of an important family of hash functions, consider the SHA-2 family, which has the same basic functional structure but varies in internal operations, message length and security bits.

A message is fed into one of these algorithms, which performs repetitive, one-way operations to produce a digest of the message. Blocks of 512-bit messages and 256-bit hash values expressed as eight 32-bit words are processed 64 times by the BHMD5 algorithm (C, D, ..., I). The hash message is 256-bits in length.

The following are some possible descriptions for the system:

$$\begin{cases} \dot{z} = l(t - z) \\ \dot{t} = xz - t - zm \\ \dot{m} = zt - vm \end{cases} \quad (12)$$

Where, z, t, and m represent the current state of the system, whereas l, v and x represent its parameters. A system enters chaos when a, b, and c are all equal to 10.

The BHMD5 was utilized to generate a 256-bit secret key J in the suggested cryptosystem. No matter how minor a difference there is between two photos, their hash values will be different. Thus, a 2256-complexity cryptosystem can withstand a brute-force attack without being breached. Accordingly, J may be represented as follows: 256-bit secret key split into 8-bit blocks (j_u).

$$J = j_1, j_2, j_3, \dots, j_{32} \quad (13)$$

The initial values can be obtained as follows.

$$z_0 = z_0' + \frac{(j_1 \oplus j_2 \oplus j_3 \oplus \dots \oplus j_{11})}{256} \quad (14)$$

$$t_0 = t_0' + \frac{(j_{12} \oplus j_{13} \oplus j_{14} \oplus \dots \oplus j_{22})}{256} \quad (15)$$

$$m_0 = m_0' + \frac{(j_{23} \oplus j_{24} \oplus j_{25} \oplus \dots \oplus j_{33})}{256} \quad (16)$$

Where, z_0', t_0' andm_0' are the initial given values.

Four elements make up the suggested encryption algorithm. Transaction data is first encrypted using the hash technique to create J and the Lorenz system's starting values. Encryption rules are then applied to the original data sequence J. Finally, decipher the series of numbers. Encrypted data is the product of the process.

Step 1: For each row and each column, enter the data in the form of O(n,b), where n is the number of rows.

Step 2: Generate the starting values of J and the key sequence (z_0', t_0',m_0') of the Lorenz system.

Step 3: The matrix N_j (n, b ×8) may be generated by repeating the binary sequence $J_{v,r}$ times where $r = \frac{n\times b\times 8}{32}$. Encode N_j with the same encoding rule and N_{jw} was obtained.

Step 4: According to XOR operation, Oe' = Oe XOR N_{jw}, Of'= OfXOR N_{je} and Ov' = OvXOR N_{jw}.

Step 5: The chaotic sequences are generated as z_b, t_bandm_b and its length is n× b× 4 by using the Lorenz system will having the initial values z_0', t_0' andm_0'.

Step 6: Prepare the chaotic sequences z_b, t_bandm_b as follows:

$$\begin{cases} (kz, dz) = sort(z) \\ (kt, dt) = sort(t) \\ (km, dm) = sort(m) \end{cases} \quad (17)$$

where (•, •) = sort (•) a new sequence after rising to an index value z is the new sequence, and the index values for this new sequence are the index values for this new sequence.

Step 7: The binary matrices should be converted $Oe', Of'$and$Ov'$to three vectors Ce (n× b× 4), Cf (n× b× 4) and Cv(n× b× 4), respectively.

$$\begin{cases} Ce'(u) = Ce(kz(u)) \\ Cf'(u) = Cf(kt(u)) \\ Cv'(u) = Cv(km(u)) \end{cases} \quad (18)$$

Step 8: Convert $\mathrm{Ce}', \mathrm{Cf}'$ and $\mathrm{Cv}'$to matrices Ew(n, b× 4), Fw(n, b× 4) and Vw(n, b×4), respectively. Decode Ew, Fw and Vw using a same rule Rule_{dec} and get three binary matrices Ev, Fv and Vv.

Step 9: Finally, recover data that is the encrypted one

Decryption is the opposite of encryption in that it is a reverse operation. In order to decode the encrypted data, the receivers must get secret keys from the sender. Following is a step-by-step breakdown of the decryption procedure.

Step 1: Having Rule_{dec}, Components of the ciphered data are encoded. Three matrices were given as a input in total. Ew, Fw and Vw, as well as their vectorization into three $\mathrm{Ce}', \mathrm{Cf}'$ and Cv'.

Step 2: $\mathrm{Ce}', \mathrm{Cf}'$ and $\mathrm{Cv}'$are the vectors. For obtaining the non- vectors Ce, Cf and Cv, step 8 of the encryption method is reversed as a result:

$$\begin{cases} Ce(u) = Ce'(kz(u)) \\ Cf(u) = Cf'(kt(u)) \\ Cm(u) = Cm'(km(u)) \end{cases} \quad (19)$$

Where, kz, kt and km It is addressed in the encryption procedure in stages 6 and 7.

Step 3: Convert the three vectors Ce, Cf and Cv in the form of the matrices Oe', Of'andOv'.

Step 4: According to XOR operation the invert process was done then the encryption algorithm as follows:Oe =Oe' XOR N_{jw}, Of = Of' XOR N_{je} and Ov = Ov' XOR N_{jw}, where N_{jw} utilizing the key sequence J, as described in encryption step 5, yields the desired result.

Step 5: Oe, Of and Ov are the three matrices for the sequence. Using the rule, the decoding was done (Oe, Of, and Ov)to extract the E, F, and V components of the data.

Step 6: Finally, the original data was recovered.

Before the data can post to the server it can be splitted to improve the second level of the security, for that the MHA was used.

Hellman Fibonacci sequence D_b can be defined as follows.

$$D_b = \begin{cases} 0 \; b = 0 \\ 1 \; b = 1 \\ D_{b-1} + D_{b-2} \; otherwise \end{cases} \quad (20)$$

Equation (20) is used to construct the Fibonacci sequence, which consists of the integers. Using any four consecutive terms of the Fibonacci numbers, a 2 x 2 matrix may be created that can be used to scramble data. The definition of a generalized Fibonacci mask is as follows:

$$\begin{bmatrix} \dot{z} \\ \dot{t} \end{bmatrix} = \begin{bmatrix} d_u & d_{u+1} \\ d_{u+2} & d_{u+3} \end{bmatrix} \begin{bmatrix} z \\ t \end{bmatrix} \bmod (b) \quad (21)$$

Where z, t, $\dot{z}$, $\dot{t}$ ∈ 0,1,2,3,4, … ,b−1 , du is the u^{th} term of the Fibonacci series, and b is the overall data size z, A crammed dataset has a new coordinate for the original data, which is t. Scan the whole data horizontally and vertically so that the data may be scrambled into new information.

Y orthogonal matrix of size n×n, C orthogonal matrix of size b×b and A diagonal matrix of size n×b are all singular value decomposition matrices for any given matrix L ∈ $E^{n\times b}$ where u≠h.

$$L_{nb} = Y_{nn} A_{nb} C_{bb}^{R} \quad (22)$$

Where $Y^T Y=I$, $C^T C=I$ and $a_{11} \geq a_{22} \geq \cdots a_{oo} \geq 0$, where o = min {n, b}. The columns of **Y** are orthonormal eigenvectors of LL^T, the columns of **C** are represented as a orthonormal eigenvectors of L^TL, where A is a diagonal matrix that contains the square root of Eigenvalues from Y or C in decreasing order. This is what matrix N would look like if it was 5 by 3.

$$N = Y \times A \times C^T \tag{23}$$

$$\begin{bmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \\ n_{31} & n_{32} & n_{33} \\ n_{41} & n_{42} & n_{43} \\ n_{51} & n_{52} & n_{53} \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & y_{13} & y_{14} & y_{15} \\ y_{21} & y_{22} & y_{23} & y_{24} & y_{25} \\ y_{31} & y_{32} & y_{33} & y_{34} & y_{35} \\ y_{41} & y_{42} & y_{43} & y_{44} & y_{45} \\ y_{51} & y_{52} & y_{53} & y_{54} & y_{55} \end{bmatrix} \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}^T \tag{24}$$

The following are the measures that will be taken:

1) First, need to take a look at the data in the form of a matrix Z.

2) Using Hellman Fibonacci Transform, randomize the elements of matrix Z to get matrix V.

3) Apply the SVD transformation to the matrix V to create three new matrices.

$$[Y, A, C] = SVD(V) \tag{25}$$

4) L and O are Diagonal matrices that contain only integer and fractional parts of A values, respectively.

$$A = L + O \tag{26}$$

5) The receiver already has the same keys (big and small) that may be utilized for the data reversing operation.

C. Equations

The attack success rate is done in a Scyther. Here$\alpha < \beta + \gamma$, the probability decreases exponentially with an increasing R_x . In this analysis, the hash power (G) will dominate the mining race in R_x time. When predicting how many blocks will be mined over the course of a certain period of time, the Poisson distribution may be used.

A double-spending assault happens when the opponent makes a second payment that is more than the previous payment. O_{sa} may be calculated by adding up all possible double spending attacks.

$$O_{so} = \sum_{J_1=0}^{+\infty}[O_1(J_1, R_x) \sum_{J_2=J_1+1}^{+\infty} O_2(J_2, R_x)], \tag{27}$$

Where, $O_1(J_1, R_x)$ and $O_2(J_2, R_x)$ are defined in Equation (27) and Equation (27), respectively.

It is the payee of the double-payment assault who has got the first payment and is trying to get a second one from Arbitrator. It is the payer's goal to defeat the payee in arbitration so that the arbitrator does not make a second payment to the latter. Hash power is used to create Non-Payment Proof in the double-payment attack βG and PaymentChallenge is generated with hash power $(\alpha + \gamma)G = (1 - \beta)G$ By the similar analysis as O_{sa}, the probability of double-payment attack O_{so} is given by.

$$O_{so} = \sum_{J_1=0}^{+\infty}[O_1(J_1, R_x) \sum_{J_2=J_1}^{+\infty}[O_2(J_2, R_x)], \tag{28}$$

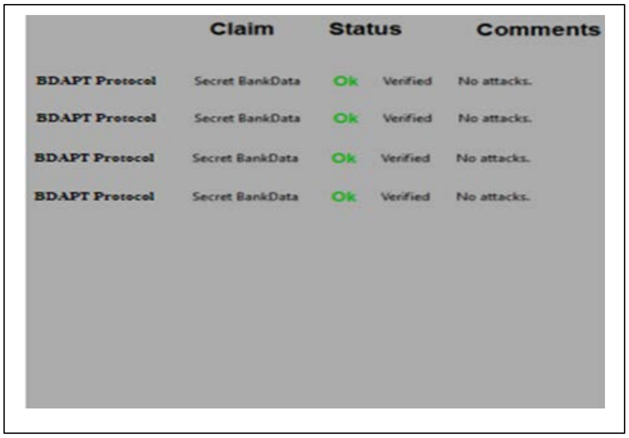
Where, $O_1(J_1, R_x)$ and $O_2(J_2, R_x)$ are dened in Equation (27) and Equation (28), respectively. If Non-Payment Proof and Payment Challenge have the same length, then Arbitrator decides in favor of the payee, and there is no other difference in the computation of O_{so} and O_{sa}. With rising R_x, O so decreases exponentially. It is therefore possible to minimize the O so double-payment chance significantly by raising the parameter R_x by using the proposed procedure.

V. Performance Analysis

Mobile payment transactions between smartphones and payment terminals will be made more secure thanks to a new protocol were introduced in this paper. It provides an extra layer of cryptographic protection to address payment security flaws. It safeguards: the integrity and confidentiality of banking information, as well as the mutual authentication and non-repudiation of those data; and the validity of those data that are not. Using the Scyther tool, the suggested methodology was tested.

Humans still have a hard time proving that a security protocol is correct and safe. So, Scyther tool was used, which has been used in both research and teaching environments, to check the suggested protocol. Scyther enables security protocols to be analyzed in a formal manner by detecting possible attacks and weaknesses. When compared to other security verification methods, the results from Scyther's researchers are impressive. Unrestricted sessions and assured end-of-life are the hallmarks of Scyther's protocol analysis service. It generates a graph describing an assault if it discovers one that matches a certain claim. Protocols in the Scyther tool are implemented using the Security Protocol Description Language (SPDL). All Scyther statements that no attacks have been detected on Banking Data are supported by the protocol, as shown in Fig. 3. Definitions based on official sources are provided below.

The proposed Logical Diffie Hellman algorithm may be shown to be successful by comparing it to the [20] in order to demonstrate its efficacy.



	Claim	Status		Comments
BDAPT Protocol	Secret BankData	Ok	Verified	No attacks.
BDAPT Protocol	Secret BankData	Ok	Verified	No attacks.
BDAPT Protocol	Secret BankData	Ok	Verified	No attacks.
BDAPT Protocol	Secret BankData	Ok	Verified	No attacks.

Fig. 3. Protocol Efficiency Analysis.

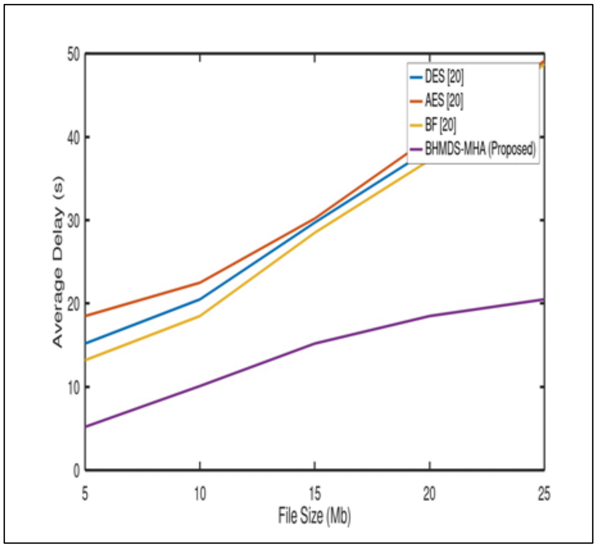


Fig. 4. File Size Vs. Average Delay.

Errors are more likely to occur during a switchover if the information used is inaccurate, imprecise, or confusing. Processing, queuing, transmission, and propagation delays are the four most common types of delays in packet switched networks. The quality and timeliness of a product are threatened by delays in communication (Transmission). Some of the probable repercussions include longer wait times, delays in discharge and poor decision-making. The delay ratio is shown in Fig. 4. The suggested method has a much lower delay ratio (20.5) than earlier procedures.

Fig. 5 and 6 depicts the time it takes to encrypt and decode a message in milliseconds, and the performance of the method is calculated. The suggested approach is compared to existing algorithms as BF, DES, and AES in terms of time differences.

It is shown in Fig. 7 that the proposed technique consumes less energy than other well-known encryption algorithms such as BF and DES.

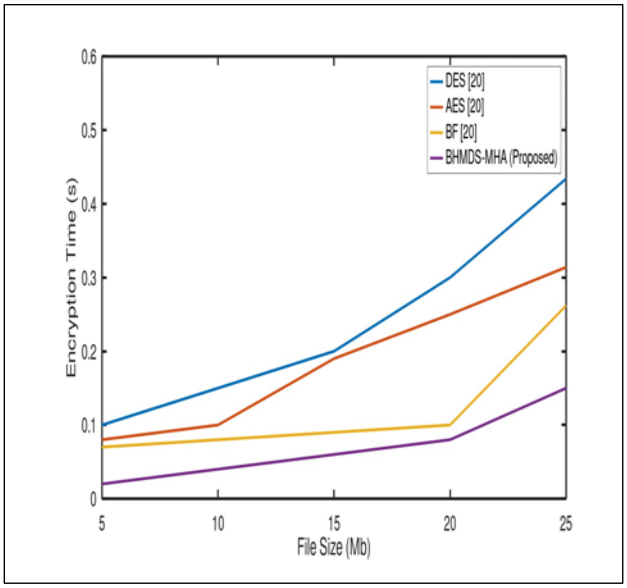


Fig. 5. File Size Vs. Encryption Time.

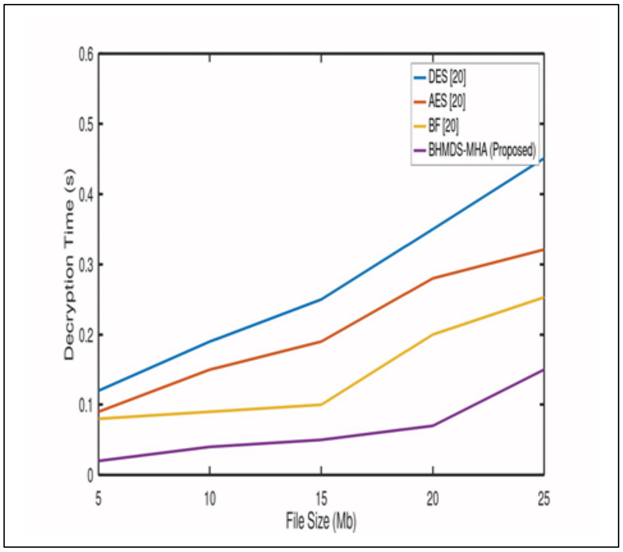


Fig. 6. File Size Vs. Decryption Time.

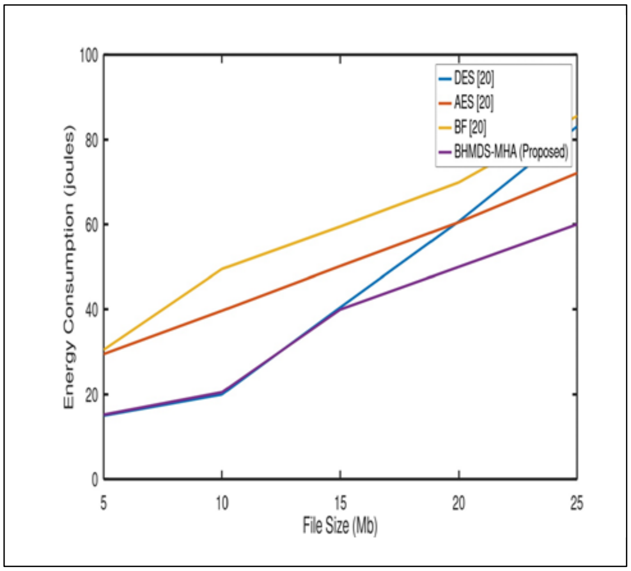


Fig. 7. File Size Vs. Energy Consumption.

Fig. 8 compares the proposed method to existing ones like DES BF and AES and shows how much faster it is in terms of throughput (in kbps). Fig. 5 to 8 show how the suggested method is tested. They look at things like average delay, throughput, encryption time, energy consumption, and decryption time to see how well the method works. BF, DES, and AES are used to figure out BHMD5-MHA, which takes into account things like average delay and throughput. BF is a competitor in near vicinity that ensures the privacy and integrity of user data. Although it seeks to reduce key complexity, its security is sometimes unreliable. BHMD5-MHA improves security for secure and successful financial transactions while also reducing the likelihood of fraud.

A secured mobile payment system was created here in this study by first employing the BHMD5-MHA method, which saves encrypted blocks of data and then links them together. An evaluation of the proposed mechanism's level of security is carried out. The proposed BHMD5-MHA algorithm is compared against existing methods as DES [18], RSA [19], and AES [21].

Various encryption algorithms are compared and contrasted in Fig. 9 to show how secure they are in comparison. The existing DES, RSA, AES methodologies are compared with a new one called BHMD5-MHA. The security level is 82% TDH, 78% DES, 77% RSA, and 73% AES for files less than 20 MB. Security is also evaluated for file sizes as large as 40 MB and as little as 1 MB. When compared to various encryption algorithms, the graph shows that the BHMD5-MHA is the most secure. According to the data, the proposed technique beats other currently used procedures when it comes to ensuring transaction security.

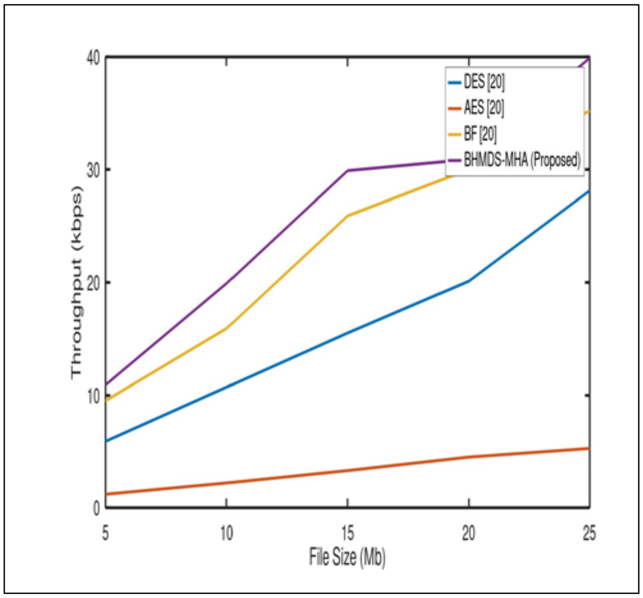


Fig. 8. File Size Vs. Throughput.

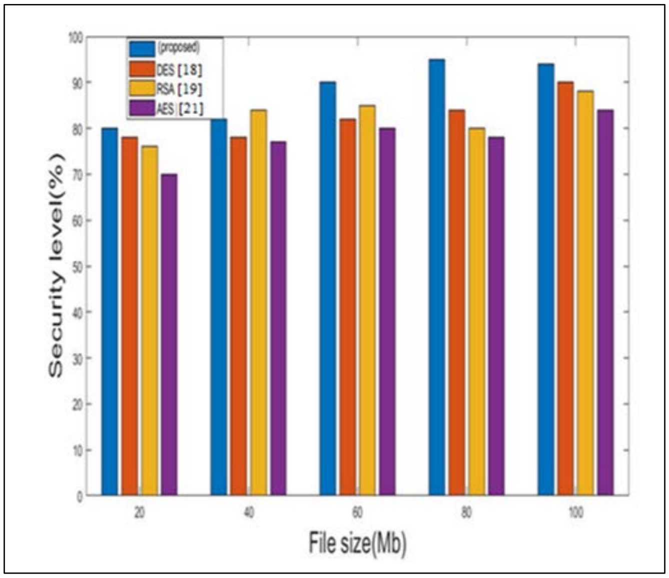


Fig. 9. File Size Vs. Security Level.

VI. Discussion

Based on the outcome that was achieved, our protocol meets the transaction security requirements listed below: 1) Symmetric encryption and the secret are used to guarantee the authenticity of the parties involved. 2) Transaction privacy is ensured by encryption, 3) Transaction integrity is ensured by the suggested protocol, and 4) Non-repudiation of transactions is also ensured in that the merchant is able to provide a non-repudiable evidence to prove to other parties that the client has originated the message. The encryption ensures that either the client or the merchant has originated the message and authenticates the client. Because the keys that are communicated between parties in our proposed protocol need to be updated on a regular basis or in response to specific requests, another worry that emerges relates to the distribution of keys. In the normal course of events, every time a new key is released, even if it is done so in an encrypted form, it is still feasible for an adversary to obtain it. Because it is possible to enlist the information in a coded form, our protocol does not need the customer to provide their card information. A participant in any transaction should not place their faith in other parties until those other parties can demonstrate that they can be trusted. Because the issuer is the one who provides the client with a credit card, our protocol specifies the trust connection that exists between the client and the issuer rather than the trust relationship that exists between the client and the payment gateway.

VII. Conclusion

The novel protocol and cryptographic techniques have been clearly addressed in this study. The Scyther compliance with standards and guidelines is used to demonstrate the model. An analogy of the key length is provided, as well as an analytical look into cryptography approaches for real-time use Due to the magnitude of the BHMD5's key, it has been demonstrated to be the most secure. As a second degree of protection, the MHA scrambles the original data. According to other symmetric key algorithms, this approach is the most secure (96 percent) and uses the least processing time. According to the findings, the proposed method could help in deal with the growing security issues that come with online transactions.

Acknowledgment

This research was supported by the Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamilnadu, India.

References

[1] F. Buccafurri and G. Lax, "Implementing disposable credit card numbers by mobile phones," Electronic Commerce Research, vol. 11, pp. 271-296, 2011.

[2] A. Braeken, "Public key versus symmetric key cryptography in client–server authentication protocols," International Journal of Information Security, vol. 21, pp. 103-114, 2022.

[3] S. Bojjagani, V. Sastry, C.-M. Chen, S. Kumari, and M. K. Khan, "Systematic survey of mobile payments, protocols, and security infrastructure," Journal of Ambient Intelligence and Humanized Computing, pp. 1-46, 2021.

[4] E. Erdin, S. Mercan, and K. Akkaya, "An evaluation of cryptocurrency payment channel networks and their privacy implications," arXiv preprint arXiv:2102.02659, 2021.

[5] H. Li, T. Wang, Z. Qiao, B. Yang, Y. Gong, J. Wang, et al., "Blockchain-based searchable encryption with efficient result verification and fair payment," Journal of Information Security and Applications, vol. 58, p. 102791, 2021.

[6] J. Zhang, Y. Ye, W. Wu, and X. Luo, "Boros: Secure and Efficient Off-Blockchain Transactions via Payment Channel Hub," IEEE Transactions on Dependable and Secure Computing, 2021.

[7] K.-H. Yeh, C. Su, J.-L. Hou, W. Chiu, and C.-M. Chen, "A robust mobile payment scheme with smart contract-based transaction repository," IEEE Access, vol. 6, pp. 59394-59404, 2018.

[8] Y. Chen, W. Xu, L. Peng, and H. Zhang, "Light-weight and privacy-preserving authentication protocol for mobile payments in the context of IoT," IEEE Access, vol. 7, pp. 15210-15221, 2019.

[9] M. Ramachandran and V. Chang, "Towards performance evaluation of cloud service providers for cloud data security," International Journal of Information Management, vol. 36, pp. 618-625, 2016.

[10] F. Corradini, L. Mostarda, and E. Scala, "ZeroMT: Multi-transfer Protocol for Enabling Privacy in Off-Chain Payments," in International Conference on Advanced Information Networking and Applications, pp. 611-623, 2022.

[11] N. El Madhoun and G. Pujolle, "Security enhancements in emv protocol for nfc mobile payment," in IEEE Trustcom/BigDataSE/ISPA, pp. 1889-1895, 2016.

[12] S. Abughazalah, K. Markantonakis, and K. Mayes, "Secure mobile payment on NFC-enabled mobile phones formally analysed using CasperFDR," in 13th International Conference on Trust, Security and Privacy in Computing and Communications, pp. 422-431, 2014.

[13] Y.-Y. Chen, M.-L. Tsai, and F.-J. Chang, "The design of secure mobile coupon mechanism with the implementation for NFC smartphones," Computers & Electrical Engineering, vol. 59, pp. 204-217, 2017.

[14] S.-W. Chen and R. Tso, "NFC-based mobile payment protocol with user anonymity," in 11th Asia Joint Conference on Information Security (AsiaJCIS), pp. 24-30, 2016.

[15] N. El Madhoun, F. Guenane, and G. Pujolle, "An online security protocol for NFC payment: Formally analyzed by the scyther tool," in Second International Conference on Mobile and Secure Services (MobiSecServ), pp. 1-7, 2016.

[16] N. El Madhoun, F. Guenane, and G. Pujolle, "A cloud-based secure authentication protocol for contactless-nfc payment," in 4th International Conference on Cloud Networking (CloudNet), pp. 328-330, 2015.

[17] J. N. Luo, M. H. Yang, and S.-Y. Huang, "An Unlinkable Anonymous Payment Scheme based on near field communication," Computers & Electrical Engineering, vol. 49, pp. 198-206, 2016.

[18] R. Shivhare, R. Shrivastava, and C. Gupta, "An Enhanced Image Encryption Technique using DES Algorithm with Random Image overlapping and Random key Generation," in International Conference on Advanced Computation and Telecommunication (ICACAT), pp. 1-9, 2018.

[19] V. Rao, N. Sandeep, A. R. Rao, and N. Niharika, "FPGA Implementation of Digital Data using RSA Algorithm," Journal of Innovation in Electronics and Communication Engineering, vol. 9, pp. 34-37, 2019.

[20] A. H. Be and R. Balasubramanian, "Encryption Algorithm for High-Speed Key Transmission Technique," Advances in Dynamical Systems and Applications, vol. 16, pp. 1557-1267, 2021.

[21] X. Dong, D. A. Randolph, and S. K. Rajanna, "Enabling privacy-preserving record linkage systems using asymmetric key cryptography," in AMIA Annual Symposium Proceedings, pp. 380, 2019.

[22] A.Saranya, R.Naresh “Cloud-Based Efficient Authentication for Mobile Payments using Key Distribution Method”, Journal of Ambient Intelligence and Humanized Computing, 2021. [Online]. Availble: http://dx.doi.org/10.1007/s12652-020-02765-7.

[23] A.Saranya, R.Naresh “Efficient mobile security for E-healthcare application in cloud for secure payment using key distribution”, Neural Processing Letters, 2021. [Online]. Availble: http://dx.doi.org/10.1007/s11063-021-10482-1.

[24] R.Naresh, P.Vijayakumar, L. Jegatha Deborah, R. Sivakumar, “A Novel Trust Model for Secure Group Communication in Distributed Computing”, Special Issue for Security and Privacy in Cloud Computing, Journal of Organizational and End User Computing, vol. 32, no. 3, 2020.

[25] R.Naresh, M.Sayeekumar, G.M.Karthick, P.Supraja, “Attribute-based hierarchical file encryption for efficient retrieval of files by DV index tree from Cloud using crossover genetic algorithm”, Soft Computing, Springer, vol.23, no. 8, pp. 2561-2574, 2019.

[26] R.Naresh, AyonGupta, Sanghamitra, “Malicious Url Detection System Using Combined Svm And Logistic Regression Model”, International Journal of Advanced Research in Engineering and Technology, vol. 10, no. 4, pp. 63-73, 2020.

[27] Gautam Srivastava, C.N.S. Vinoth Kumar, V Kavitha, N Parthiban, Revathi Venkataraman, “Two-Stage Data Encryption using Chaotic Neural Networks”, Journal of Intelligent and Fuzzy Systems, vol. 38, no. 3, pp. 2561-2568, 2020.

[28] R. Mugesh, “A Survey on Security Risks in Internet of Things (IoT) Environment,” Journal of Computational Science and Intelligent Technologies, vol. 1, no. 2 pp. 01-08, 2020. [Online]. Available: https://doi.org/ 10.53409/mnaa.jcsit20201201.

[29] R. Sathiyasheelan, “A Survey on Cloud Computing for Information Storing,” Journal of Computational Science and Intelligent Technologies, vol. 1, no. 2 pp. 09-14, 2020. [Online]. Available: https://doi.org/ 10.53409/mnaa.jcsit20201202.

[30] A.N. Suresh, “A Hybrid Genetic-Neuro Algorithm for Cloud Intrusion Detection System,” Journal of Computational Science and Intelligent Technologies, vol. 1, no. 2 pp. 15-25, 2020. [Online]. Available: https://doi.org/ 10.53409/mnaa.jcsit20201203.

[31] C.N.S.Vinoth Kumar, and A.Suhasini, “Secured Three-Tier Architecture for Wireless Sensor Networks Using Chaotic Neural Networks”, Advances in Intelligent Systems and Computing, vol. 507, no. 13, pp. No. 129-136, 2017. [Online]. Available: https://doi.org/10.1007/978-981-10-2471-9_13.

[32] C.N.S.Vinoth Kumar, and A.Suhasini, “Improved secure three-tier architecture for WSN using hop-field chaotic neural network with two stage encryption,” in International Conference on Computer, Electrical & Communication Engineering 2016. [Online]. Available: https://doi.org/*10.1109/ICCECE.2016.8009540*.

An Efficient Approach towards Vehicle Number Estimation with Ad-hoc Network under Vehicular Environment

Yuva Siddhartha Boyapati[1], Shallaja Salagrama[2]
Ph.D Scholar, Doctor of Philosophy, Information Technology, University of the Cumberland's
Williamsburg, Kentucky, USA

Vimal Bibhu[3]
Associate Professor, Department of Computer Science & Engineering, Amity University Greater Noida Campus
Greater Noida, India

***Abstract*—Ad-hoc network usability extends the application for Dedicated Short-Range Communication. This type of ad-hoc network technology is non infrastructure and due to this fact, it can be used for Direct Short Range Communication System to provide the quick and real time message to the vehicular operator to prevent the damage and fatality of life by meeting accidents and crashes. In this paper, we present a holistic approach to estimate the number of vehicles in specified range of one KM distance. The designed system for vehicle number estimation is based on the Time Division Multiple Access mechanism which further estimates the number of reserved slots by vehicular nodes. This estimation methodology is tested under the digital simulator and approximately 34 number of vehicles for 24 seconds are defined to test the slot reservation. We found that in case of vehicular nodes greater than 20, slot reservation accuracy is 95% and when the vehicular nodes are less than 20 then the slot reservation is 100%.**

Keywords—Vehicular ad hoc networking; hidden vehicle; visible vehicle; time division multiple access; dedicated short range communication

I. Introduction

Dedicated short range communication (DSRC) structure was specially developed for the vehicular communication technology. A bunch of wireless methodologies are available for the primary medium of communication to DSRC. The major needs of wireless system under ad-hoc network for vehicular environment is that the latency should be 20 mili seconds or less than it, throughput should be high, and it cover the communication range up to one kilometer. There should also be support of diverse schemes of communication with the wireless technology. The important one is support of one-way in both uplink and downlink communication allowing the vehicular nodes to dissipate a broadcast message. The second important communication requirement is two-way that allows two vehicles to establish the dialog between them. Third important one communication requirement is point to point and point to multipoint. The point to point communication supports the message delivery to specific location or vehicular node, and point to multipoint communication the message is sent to many locations or vehicular nodes.

It is much important to evaluate the wireless technologies to determine which particular meets the DSRC requirements. A modified version of IEEE 802.11a with Time Division Multiple Access (TDMA) methodology equipped Physical layer is considered to be best suitable for DSRC communication [1]. Also, the other evaluated wireless technologies are not suitable so that these technologies are unacceptable with many different reasons. The cellular and satellite communication systems provide huge bandwidth but these two technologies have high latency. Due to this fact cellular and satellite technologies are suitable only for some of applications of DSRC. The cellular technology does not support the message broadcasting and it is a costly communication system. Similarly, satellite communication system is costly and is not suitable for DSRC applications. The wireless technology under the DSRC applications should be free of cost and must be based on ad-hoc communication. Also, the cost pertaining to the infrastructure must be cheaper than that of the satellite and cellular communication system.

Vehicular ad hoc network (VANET) becomes a challenging task to implement in real world scenario. Ad-hoc network in vehicular environment is a type of mobile ad hoc network where the computing nodes are being implemented on vehicles with modern technique of communication devices. The communication of information from vehicle to vehicle or vehicle to roadside and vice versa is performed with the active help of fast computing environment. The clear scenario for feasible vehicular communication for the different categories of applications is either related to public safety or general, such as internet accessibility based upon the various factors of vehicular ad hoc networking. One of the major problems related to estimate the number of vehicles in a lane with respect to the coverage area of an antenna established at roadside. The roadside installed antenna is also equipped with modern fast computing devices. The major role of roadside installed station is to provide the information related to vehicle safety along with structural aspect of road and highway. The estimation of number of vehicles in the specified range of the road or highway is very challenging task due to dynamic nature of vehicles.

II. Literature Review

Different forwarding methods based on the broadcasting are infrastructure less node to node communication. It is facilitated by ad hoc network under vehicular environment. The broadcast contention control provides the reliable and low latency multi-hop connection. Algorithm behind the broadcast contention control optimizes the back off distribution and provides the priority to forward the information related to the location of the node [2].

Ad-Hoc network under the vehicular environment provides the connection among the vehicular nodes of vehicles on the road. This is considered an important and valuable concept to improve the safety of the transportation system with elegant efficiency. Time division multiple access relies on slotted frame structure having the good quality of service framework. This has low scalability and complex synchronization mechanism. The vehicle number estimation according to the slot reservation is not always be optimal due to the delay and less quality of service having interference and the speeds of the nodes running on the vehicle [3].

Ad-Hoc network under the vehicular environment is more vulnerable to the Denial-of-Service attack [4]. One the most common Denial of Service Attack is Jamming attack which leads to the interference with the used wireless ad-hoc network service and creates the network congestion. Once the Jamming attack progresses, the quality of service of the ad hoc network degrades and the optimal service cannot be achieved. Due to this fact the vehicle number estimation fails and this causes the failure of the overall network service of ad hoc network in vehicular environment.

A cross layered Medium Access Control having clustering mechanism supports the propagation of the broadcast message in ad hoc network under vehicular environment [5]. Distributed dynamic clustering frames, and the dynamic virtual backbone in the ad hoc network are the building blocks of the cross layered Medium Access Control. The members associated and connected with the ad-hoc network in vehicular environment are responsible to implement the efficient message propagation to utilize the slot and count the numbers of allocated slots in the vehicle number estimation. This fails when the vehicle's speed goes beyond the defined threshold value and the performance of this method degrades.

III. Proposed Method for Vehicle Numbers Estimation

In case of vehicular ad hoc networking there are two different categories of vehicles on road or highways. These two categories of vehicles are either visible or hidden vehicles. To estimate the number of vehicles, it should be first determine the estimation method for both hidden and visible vehicles. The hidden vehicle is considered hidden because this type of vehicles has not joined the coverage area of roadside antenna. The estimation of visible or hidden vehicles are taken with the help of reservation of slots under Time division multiple access (TDMA).

A. Visible Vehicle Numbers Estimation

The intensity of vehicles with respect to specified time in a given direction is estimated by the number of occupied slots of TDMA channel by sequence numbers from beginning to end [6]. The specific beacon frame containing message indicates by adding an additional field into the frame format. This estimation of number of vehicles is performed with every regular interval of time by roadside station with broadcasting to all vehicles in the coverage range of station. These visible vehicles become under the control of regulation. The determination of visible vehicles is shown with Fig. 1.

In accordance with Fig. 1, it is very clear that the road is only two-lane structure. For left side running vehicles in a lane under same direction reserves the slots of TDMA from left to right [7]. The vehicles running into right side in other lane under opposite direction on road reserve to slot of TDMA in reverse direction. The sequence numbers of reserved slots are not repeated in any case. Hence, the sum of highest sequence numbers of both lanes TDMA slots forward and reverse directions gives the total number of visible vehicles on the coverage area of roadside station.

B. Hidden Vehicles Number Estimation

Roadside installed stations checks the number of hidden vehicles with regular interval of time by beaconing itself. The beacons with additional parameters are listening to all the vehicles but vehicles avoid this which is already connected into the network with specified group. An additional sub field that contains the connecting request information for the vehicle node is attached under the beacon frame. There is specified number of two slots under the TDMA slots which depict the beaconing for estimating the hidden vehicles. These two slots cannot be probed and reserved by any vehicles in any direction. These slots are indicated by special identifier Hidden Vehicles (HV) [8]. Under this slot interval any vehicle probing is observed then it is concluded that this is hidden vehicles and roadside unit add one in hidden vehicles number and subsequently broadcast the message to all groups about the hidden vehicle. The TDMA slots are reserved for HV predicted under Fig. 2.

The TDMA channels slots of Fig. 2, shows that the first HV is in a forward direction and second in the reverse direction of vehicles.

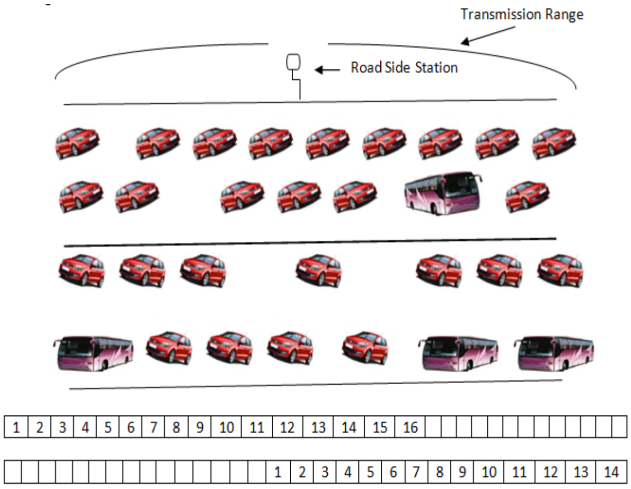


Fig. 1. Visible Vehicles having Ad-Hoc Network Node.

1	2	3	4	5	6	8	9	10	HV	
HV	1	2		3	4	5	6	7	8	9

Fig. 2. TDMA Slots with Hidden Vehicles.

Suppose that the probability is p of hidden vehicle which transmit HV frames under the TDMA slots and n is the total number of hidden vehicles in both directions of vehicles in two lanes of given road. Again, let, n1 be the number of vehicles in one direction and n2 be the number of vehicles in opposite direction. Also, let m be the total number of hidden vehicles under the communication range of a roadside station, we can derive the probabilities of the different status of HV Channel slots.

Probability that HV slot is ideal

$I = (1 - p)^{m}$

Probability that HV frame is successfully received from a hidden vehicle in a direction is

$DP_1 = n_1p(1 - p)^{m-1}$

Probability that HV frame is successfully received in opposite direction hidden vehicle is

$DP_2 = n_2p\ (1 - p)^{m-1}$

Therefore, the total number of estimated hidden vehicles in each same and opposite directions are calculated as,

$$n1 = \frac{s1}{I}\left(\frac{1-p}{p}\right)$$

$$n2 = \frac{s2}{I}\left(\frac{1-p}{p}\right)$$

After getting the probabilities values of vehicles in each direction we can further smooth the values by exponential moving averages by smoothing factor VHAlpha, which is a system parameter [9].

IV. Designed System Simulation

The estimation of hidden and visible vehicles simulation is performed with the above-mentioned criteria with total number of 30 vehicular nodes. The vehicles that are under the vehicular ad-hoc network environment must have the digital map under the designed simulator application. We have used constructed files as a digital map to simulate the VANET. The constructed files contain detailed geographical information about the road. The data which digital map contains comes in the form of geographical coordinates (latitude, longitude) for the road [10]. The constructed files based on the selected road specify the end points, having the intermediate points those are required. In case, when the road is straight, then there is not any intermediate points. This is so because the intermediary points are calculated by the help of interpolation. Further, in case of the road having curves, there are very huge numbers of intermediate points, so that the map is considered accurate.

The traffic simulator that takes actions of each of the vehicle under the given scenario is a digital simulator, as opposed to those digital simulators that use the global measure like density of traffic on road to describe the evolution of traffic. The dynamics of the traffic can be better understood with help of the digital simulator and appropriate design of facilities related to road traffic such as lanes, lights, closure of lanes and corners [10].

A. Digital Traffic Simulator

A digital traffic simulator is an application used to simulate the traffic behavior with respect to the wireless ad-hoc network communication message. Each of the vehicles which are taken under the vehicle has the computing node and is able to send and receive the wireless signals to the others. Simulating in digital traffic simulator there is requirement of setup of the vehicles and road with trajectories and intermediary point. The vehicle behavior with respect to the requirements to fulfill the order of traffic is observed in the process of simulation.

With above facts, there is requirement of high level of details to study the suitability of wireless ad-hoc network in vehicular environment. To solve this, we have designed a digital traffic simulator based on a model of behavior of driver. This designed driver behavior model is presented in Fig. 3. There are many traffic simulators such as ‘VISSIM’ and others that use this type of designed driver behavior model for the simulation of vehicular environment. Basically, VISSIM’s purpose is to model and forecast the traffic flow of vehicles, addition of new lane, making of the study of traffic closure to lane and creating the overpass and many more. Also, it is true that this type simulators are very difficult to integrate with the network simulator as these types of simulators are commercial products [12] [13].

Thereafter, we have made the description of the driver behavior model which we have deployed. The assumptions are taken under the driver behavior model are ‘free driving’, ‘following’, ‘braking’, and ‘approaching’.

Driver model with ‘free driving’ states that there is no influence by preceding vehicles under the same lane of the vehicles. Thus, the drivers have to maintain the desired speed. Although, the speed and acceleration depend on the driver and the features of the road on which the vehicles are driven.

Driver model with ‘approaching’ states that the preceding vehicle is slower in speed. In this case the driver has to de-accelerate the speed to make the same speed of preceding vehicle. Basically, de-acceleration is treated as a function of distance between preceding and approaching vehicles, their speed and other parameters in same lane of road.

Driver model with ‘following’ states that the speeds of preceding and following vehicles are equal or same. In this case the drivers of both of the vehicles maintains the same speed and constant acceleration can be maintained.

Driver model with ‘braking’ states that the slower preceding vehicle is close to front of following vehicle. In this case, the following vehicle driver has to deaccelerate by braking the vehicle to avoid the crash.

The rules defined for driver to determine the driver mode is presented by Fig. 2. Two thresholds are taken, where ‘distance1’ is first and ‘distance2’ is second, when the preceding vehicle is closer to ‘distance1’ and slower than the just following vehicle.

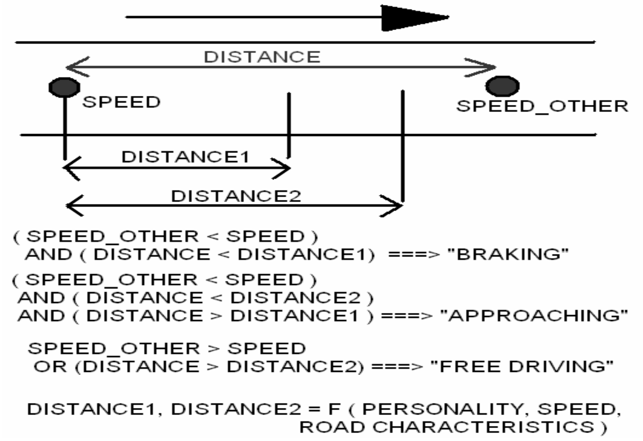

DISTANCE
SPEED
SPEED_OTHER
DISTANCE1
DISTANCE2
(SPEED_OTHER < SPEED)
AND (DISTANCE < DISTANCE1) ===> "BRAKING"
(SPEED_OTHER < SPEED)
AND (DISTANCE < DISTANCE2)
AND (DISTANCE > DISTANCE1) ===> "APPROACHING"
SPEED_OTHER > SPEED
OR (DISTANCE > DISTANCE2) ===> "FREE DRIVING"
DISTANCE1, DISTANCE2 = F (PERSONALITY, SPEED, ROAD CHARACTERISTICS)

Fig. 3. Designed Digital Traffic Simulator.

In case of ‘braking’ mode, when the slower preceding vehicle is under the ‘distance1’ and ‘distance2’, then the mode is considered ‘approaching’ then the current vehicle definitely deaccelerates [14]. When the preceding vehicle is away from distance2, then there is no influence to current vehicle by any means, and it is taken under ‘free driving’ mode. Also, the defined thresholds ‘distance1’ and ‘distance2’ are not fixed. These depend on the driver driving style and speed of the vehicle [15] [16].

B. Simulation Data

The simulation data is based on the number of vehicles taken under simulator, lanes defined under the simulator, time specified for simulation and vehicle flow per second. The whole data is presented in Table I.

TABLE I. SIMULATION DATA

No. of Vehicles	No. of Lanes	Time (Sec.)	TDMA Slot Reserved
27	2	24	25
26	2	24	24
28	2	24	25
28	2	24	25
27	2	24	25
29	2	24	27
25	2	24	24
13	2	11	13
12	2	10	12
11	2	11	11
12	2	11	11
32	2	21	31
33	2	22	31
32	2	20	20
34	2	20	32
33	2	21	31
34	2	22	32

C. Simulation Result

The scenario of simulation provides the estimation of vehicles having node connection with station and slot reservation under the range within 24 seconds. The result shows that maximum 5% deviation in case of the heavy traffic (No. of vehicles 30 or more), and in case of the vehicular nodes less than 20 the vehicle estimation is 100% in both visible and non-visible scenarios. The simulation result is presented in Fig. 4.

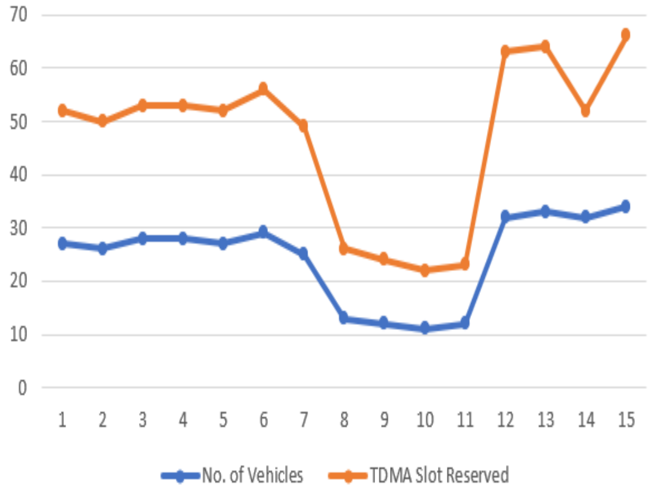


Fig. 4. Simulation Result of Vehicles Number Estimation.

The mentioned data is gathered from real time environment of the roads having two lanes under the simulator to simulate the performance. According to the simulation setup as per the given parameters the time is obtained. The TDMA slot reservation is based on the vehicles on both of the lanes under the simulation environment. Finally, the result of the simulation under the Fig. 4, is well presented.

V. CONCLUSION

The carried research work is performed to enhance safety of the living beings while travelling via vehicles like car, bus and others by employing the ad-hoc network and nodes on the vehicles. In this research work we have done the design of the digital traffic simulator and used various parameters such as approaching, distance to generate the warning message to the operators through the ad hoc network nodes to driver screen. TDMA slot allocation is used to estimate the total number of vehicles in the pre-defined range of road having single and double lane in the city scenario. The number of vehicle estimation and its results show that during the more load of vehicles, means vehicle nodes > 20 the correctness is 95% for estimation and if the number of vehicles having ad -hoc nodes < 20 then correctness is about to 100%

REFERENCES

[1] Jahanzeb Farooq, Bilal Rauf “Implementation and Evaluation of IEEE 802.11e Wireless LAN in GloMoSim” Department of Computing Science Umeå University Sweden.

[2] Fei Ye, Raymond Yim, Jianlin Guo, Jinyun Zhang and Sumit Roy, “Prioritized Broadcast Contention Control in VANET”, IEEE International Conference on Communications (ICC).

[3] Antonella Molinaro and Claudia Campolo, "MAC layer design in Vehicular Ad Hoc Networks: challenges, solutions, and methodologies", in: R. A. Santos, A. E. Block, V. R. Licea, Wireless Technologies in Vehicular Ad Hoc Networks: Present and Future Challenges, Ed. IGI Global.

[4] Ali Hamieh, Jalel Ben-Othman and Lynda Mokdad," Detection of Radio Interference Attacks in VANET", IEEE, Global Telecommunications Conference, 2009, pp: 1-5, GLOBCOM 2009 IEEE.

[5] L.Bononi, M. Di Felice, "A Cross Layered MAC and Clustering Scheme for Efficient Broadcast in VANETs", tech., http://www.cs.unibo.it/~bononi/.

[6] Giancarlo Fortino, Alfredo Garro, Samuele Mascillaro, Wilma Russo "Modeling Multi-Agent Systems through Event-driven Lightweight DSC-based Agents" Università della Calabria, Via P. Bucci cubo 41c 87036 Rende (CS) Italy.

[7] Shie-Yuan Wang, Chih-Che Lin "NCTUns 5.0: A Network Simulator for IEEE 802.11(p) and 1609 Wireless Vehicular Network" Department of Computer Science National Chiao Tung University Hsinchu, Taiwan.

[8] S.Y. Wang, C.L. Chou, C.H. Huang, C.C. Hwang, Z.M. Yang, C.C. Chiou, and C.C. Lin "The Design and Implementation of the NCTUns 1.0 Network Simulator" Department of Computer Science and Information Engineering National Chiao Tung University, Hsinchu, Taiwan.

[9] Christoph Sommer and Falko Dressler, "The DYMO Routing Protocol in VANET Scenarios",IEEE 66th Vehicular Technology Conference, 2007, VTC-2007 Fall, page 16-20.

[10] R. Mangharam, D. S. Weller, D. D. Stancil, R. Rajkumar, and J. S. Parikh, "GrooveSim: A topography-accurate simulator for geographic routing in vehicular networks," in Proceedings of the 2nd ACM International Workshop on Vehicular Ad Hoc Networks, Cologne, Germany.

[11] Bononi L, Di Felice M, Bertini M and Croci E 2006 Parallel and Distributed Simulation of Wireless Vehicular Ad Hoc Networks. Proc. of the ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '06), pp. 28–35, Torremolinos, Spain.

[12] Jérôme Härri, Marco Fiore "VanetMobiSim – Vehicular Ad hoc Network mobility extension to the CanuMobiSim framework" Institut Eurécom Department of Mobile Commu 06904 SophiaAntipolis, France Politecnico di Torino Dipartimento di Elettronica Corso Duca degli Abruzzi 24, Torino, Italy.

[13] VISSIM. http://www.vissim.de/index.php?id=1801. Accessed 9th February, 2022.

[14] Chen Q, Schmidt-Eisenlohr F, Jiang D, Torrent-Moreno M, Delgrossi L and Hartenstein H 2008 Overhaul of IEEE 802.11 Modeling and Simulation in NS-2. Proc. of the ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM '08), Chania, Greece. Code part of the official NS-2.33 release.

[15] Wu H, Lee J,HunterM, Fujimoto RM, Guensler RL and Ko J 2005 Simulated Vehicle-to-Vehicle Message Propagation Efficiency on Atlanta's I-75 Corridor. Transportation Research Board Conference, pp. 82–89.

[16] Zhou B, Xu K and Gerla M 2004 Group and Swarm Mobility Models for Ad Hoc Network Scenarios using Virtual Tracks. Proc. of the IEEE Military Communications Conference (MILCOM '04), pp. 289–294.

Virtual Tourism and Digital Heritage: An Analysis of VR/AR Technologies and Applications

Muhammad Shoaib Siddiqui, Toqeer Ali Syed, Adnan Nadeem, Waqas Nawaz, Ahmad Alkhodre
Faculty of Computer and Information Systems, Islamic University of Madinah
Madinah, Kingdom of Saudi Arabia

***Abstract*—During the time of the pandemic, travel restrictions have impacted the tourism industry with an estimated loss of more than a trillion USD; however, at the same time, we have seen a significant increase in profits for the industries which empower remote connectivity. Various studies have identified the positive impact of virtual tourism, in which tourists can be attracted by providing a VR/AR-based experience of the destination. Similarly, virtual, mixed, and augmented realities are being used to enhance user experience in digital heritage and its preservation. With emerging technologies and increasing demand for e-tourism (due to travel restrictions), there is a need to review the technological changes and analyze user requirements with respect to virtual tourism. This paper provides a literature review of the latest technologies and applications that can potentially benefit the virtual tourism and digital heritage industry. We also provide an analysis of its impact on user experience, awareness, and interest, as well as the pros and cons of virtual experiences, which may benefit the research community about the current spectrum of virtual tourism and digital heritage.**

Keywords—Virtual tourism; digital heritage; virtual reality; user experience

I. Introduction

Virtual Reality (VR) [1] is the art of envisaging the imagination capabilities of the human mind. People can visualize and interact with imaginary objects in a fantasy environment. They can pretend to live, roam, and interact in ideal, purpose-built, or inventive worlds. In general, it is non-trivial to define limits on human imagination, however, it is possible to create human-inspired imaginary worlds with the state-of-the-art technologies in computer-based systems [2]. With the advent of technology, the imagination of the virtual world can be diversified with the real world to provide more realistic experiences. This field is known as Mixed Reality (MR) [1], in which a mixture of real world and virtual objects and scenes can be emulated to provide a better and more realistic experience. Similarly, the real world can be augmented with virtual assets and entities for providing past and future experiences, in-depth infographics, and supplementary details. This type of environment is termed as Augmented Reality (AR) [3].

During the pandemic, most countries posted a ban on traveling, which had an adverse effect on the tourism industry [4]. The tourism industry was estimated to be USD 1.478 billion-dollar industry worldwide [5] before the pandemic; however, due to the COVID restrictions, the severe effects were not limited to internal and external tourism but also propagated to other related industries such as hotels, restaurants, theme parks, travel agencies, etc. Tourism and related industries are estimated to observe a Trillion USD loss in 2021 and 2022 [6].

Virtual Tourism (VT) provides ample opportunity and a substitute for the tourism industry. For the safety of the public, people were advised to keep a distance from each other, which introduced boring and monotonous experiences in everyday life. VT enables a tourism experience within the boundaries and guidelines of the pandemic restrictions. Users can have an immersive experience of a virtual or a real environment using special equipment from the comfortability of their homes. Although the user interaction in a virtual environment is not real and limited, i.e., without the experience of culture and language; however, it provides a cost-effective solution with high-quality graphics and sufficient control on viewing details in a safe and secure environment. Another advantage of virtual tourism is the preservation of cultural and natural heritage from vulnerable environments, due to tourist visits and urbanization, which introduces a potential application of Digital Heritage (DH) [7]. Digital heritage is the use of digital media to understand and preserve cultural or natural heritage. The cultural, historical, scientific, educational, and linguistic resources are converted into digital media to be preserved for later generations. For example, if a historical site is visited more often by many tourists with poor maintenance, it may deteriorate its originality. Similarly, if an ancient book, is given physical access to many people, then the vital piece of history may get destroyed. Therefore, access to these resources should only be through virtual, augmented, and/or mixed realities to provide an immersive experience.

Virtual tourism and digital heritage are enabled by VR/AR/MR technologies and applications. Without the implementation and support of the right tools, this promising industry cannot develop and thrive. In this paper, we have reviewed various technologies and applications related to virtual, augmented, and mixed realities that play a vital role in VT and DH. We have conducted a survey analysis of the perceptions, experiences, and intentions of users of different age groups about virtual tourism and digital heritage. We have also presented an in-depth analysis of commercial technologies and web/mobile applications that would benefit the research community about the current spectrum of virtual tourism and digital heritage.

The rest of the paper is structured as follows. Section II provides a detailed introduction to virtual tourism, and digital heritage and its business value. Section III reviews the

This work is supported by the Ministry of Education under project 20/16, titled: “Virtual journey into the history of the Kingdom of Saudi Arabia”.

concepts, technologies and applications of VR/AR/MR. Section IV presents the analysis of our conducted survey about peoples' perceptions of VR & DH. Section V analyzes the pros and cons of VR & DH. Section VI provides the literature review of the research related to our domain and Section VII concludes the paper.

II. Virtual Tourism and Digital Heritage

The user experience in virtual, augmented, and mixed environments can be classified into non-immersive, semi-immersive, and fully immersive experiences. Due to the everyday interaction with the artificially generated and enhanced media, people do not consider the non-immersive experience as VR/AR/MR. People are accustomed to TV, mobile phones, and computer screens without realizing that they are having a non-immersive experience in the context of VR/AR/MR. This technology provides a computer-generated experience, but still gives users awareness and control of their physical environment. Non-immersive virtual reality systems rely on computer or video game consoles, displays, and input devices such as keyboards, locators, and controllers. Watching movies and playing video games are some examples of a non-immersive VR experience [8].

The semi-immersive environment provides users a more perceptive experience of being connected to the virtual world; however, they can still feel their physical surroundings. Semi-immersive realities use high-end computer graphics to provide a more realistic environment with limited interaction, such as, using tools and simulating to have a hands-on experience using controllers and feedback actuators. The applications of semi-immersive reality are mostly simulations, which partially replicate the design and functionality of real-world mechanisms, for training and educational purposes [8].

The most realistic and attractive experience is provided by the fully immersive realities. It provides the sense of touch using hand gestures and feedback actuators, sense of sound by using stereo-based and motion-based directional audio, and sense of sight by using motion and eye tracking sensors. Head-mounted 3D displays with directional sensors are used to provide a wide field of 360-degree stereoscopic view. This type of VR has generally been adapted for gaming and other entertainment purposes, but its use is also increasing in other fields such as education, training, and virtual tourism. With fully immersive experience, the potential for using VR/AR/MR is endless and only limited by human imagination [8].

All three categories of VR/AR/MR experience can be used to enable virtual tourism and digital heritage; however, they are limited by cost-effectiveness, in-depth design, technological evolution, and user acceptance [9].

A. *Virtual Tourism*

Virtual tourism is a remote application that enables travelers to explore nature, attractions, destinations, ruins, buildings, and other travel destinations without having to physically visit them. It provides a detailed experience made possible by the emerging technology. Virtual tours can include 360-degree photos, guided teleconferencing, virtual reality (VR), augmented reality (AR), video tours, the ability to interact with art and culture experiences, or ancient environments [10].

Virtual tourism is a growing trend around the world and it's not just a response to the COVID epidemic, it is being developed behind the scenes for quite some time. Traditionally used primarily as a marketing tool, virtual tourism has recently become increasingly popular among tourism industry stakeholders. Thanks to technological advances around the world and the use of the Internet, it is deeply rooted in the concept of smart tourism. Worldwide, the VT industry was worth USD 5 Billion in 2021 and it is anticipated to reach USD 24 Billion in 2022 [11], as shown in Fig. 1.

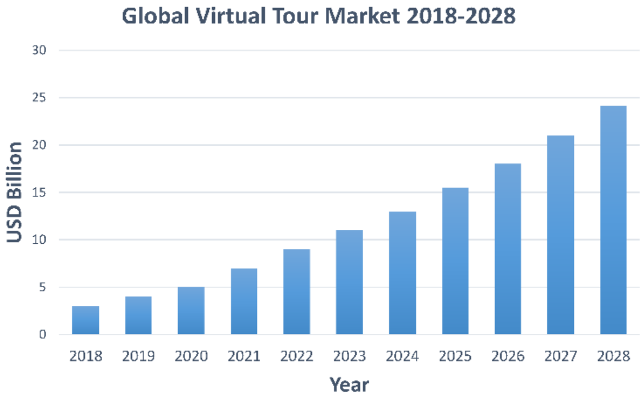


Fig. 1. Market Growth of Virtual Tourism [11].

Virtual tourism takes many forms and offers different levels of technical competence. In its simplest form, virtual tours solely rely on videos of places of interest. A 'tourist' uses speakers and a screen to see pictures. A more sophisticated form of virtual tourism involves immersing yourself in the environment using a headset or simulator. This may involve the use of various technological instruments, the user may need to wear gloves, and may have additional sensors, such as motion, proximity, direction, gestures, feeling (response), and even smell. Virtual tourism covers a wide range of digitally enhanced realities, including virtual reality, mixed reality, and augmented reality. For instance, whether you are a traveler looking for an exotic destination like Forest of Knives in Madagascar, a teacher sharing an experience with your students about the ruins of Machu Picchu, an architectural enthusiast looking for modern design like the Guggenheim in New York, or a wildlife explorer to spot animals from a safari jeep in Tanzania at sunrise, there is a virtual tour available for you.

Moreover, the virtual museum tour at the Louvre [12] provides access to Egyptian Antiquities, the remains of the Louvre's Moat, the Galerie d'Apollon, and famous artworks, such as, the Winged Victory of Samothrace, Mona Lisa, the Coronation of Napoleon, and many more. Dubai 360 provides an immersive awe-inspiring view of Dubai's modern architecture [13]. An interactive journey to the Great Barrier Reef, narrated by David Attenborough is available on YouTube 3D [14]. Users can have a semi-immersive experience of Anne Frank's house, walk through the Great Wall of China [15], indulge in the royal history at Buckingham Palace [16], experience a flight over a Volcano, bewildered by the scenic view of the Yosemite National park [17], imagine the gladiators' battle inside the Colosseum [18], inspired by the 360 view of the Statue of Liberty [19], visit the Holy land of Jerusalem [20] and Mecca [21], climb Mount Everest [22], get lost in the Amazon jungle, or even take a trip to space [23]. From the couch of your living room, you can party in Ibiza, be alone in the clouds in a hot air balloon, dive with fishes in the Georgia aquarium, or fly over the skies of Paris; virtual tourism enables all of this and much more [24].

B. *Digital Heritage*

The Charter on the Preservation of Digital Heritage of UNESCO defines digital heritage as embracing "cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources" [25]. UNESCO divides digital heritage into two separate groups namely Digital Cultural Heritage and Digital Natural Heritage [25]. Cultural heritage is the preservation and recording of cultural entities through digitization. These are things that are considered culturally important, which can be digitized or presented in physical details. It also includes intangible heritage, such as, traditions, customs, value systems, techniques, traditional dances, food, performances, and other unique features of culture, which are prone to destruction due to urbanization. Digital natural heritage is related to natural heritage objects of cultural, scientific, or aesthetic importance. In this case, digital heritage is used not only to provide access to these objects, but also to monitor changes such as plant and animal habitats over time.

There are several projects and programs which concentrate on digital heritage [27]. Global Digital Heritage (GDH) is a non-profit, private research and educational organization dedicated to documenting, monitoring, and preserving the global cultural and natural heritage [28]. They record, manage and provide immersive experiences to various destinations like Château de La Roche-Guyon in France, Early Christian Basilica of Sofiana in Italy, Archaeological Museum of Aidone in Italy; Bull Ring and Sanctuary of the Virgen de Las Nieves in Spain, Necropolis of Jebel al-Buhais in UAE, and many more [26].

Although a tour to the museum brings a surreal sensation of being in the past and present at the same time, feeling interested, surprised, happy, fantastic, and special because one can learn new things about the history and the art. Google Arts & Culture works with more than 1,200 museums and galleries around the world to provide everyone with virtual tours and online exhibits of the world's most popular museums [29]. Preserving the heritage by digitizing and making it available to the people is the main goal of such initiatives. Google Arts and Culture do not include all the museums and galleries; however, many of the museums, cultural sites, galleries, and even literature sites have created 2D and/or 3D models, which can be experienced through either virtual reality or augmented reality.

Similarly, history enthusiasts are able to visit various sites (Machu Picchu in Peru, stone carved palaces of Petra in Jordan, Taj Mahal in India, Khumbu Valley in Nepal, the Vatican City, etc.) [29], museums (Smithsonian National Museum of Natural History in USA, British Museum, etc.) [30] and read digitized old books at libraries (Library of Congress, British Library and UC Davis Library, etc.) [31] without disturbing the physical environment.

III. VR/AR Technologies and Applications

Virtual Tourism and Digital Heritage are enabled by VR/AR/MR technologies and applications [32] [33]. Without the implementation and support of the proper tools, this promising industry cannot evolve and prosper [34]. Various technologies are required for creating a virtual world, which involves capturing the real world, modelling the assets and scenes, and adding narration and interaction. For experiencing the virtual world, involves immersive displays, movement/motion tracking, environment sensing, and enabling runtime command and control. These technologies can be classified as, display screens, viewing glasses, sensors, cameras, image processing tools, and graphics processors. Furthermore, various software are also required to provide realistic experience, such as graphics engines, AR/VR toolkits, and libraries and APIs. In the following subsections, the models that envisages VT and DH, and some of the enabling technologies and application toolkits are discussed.

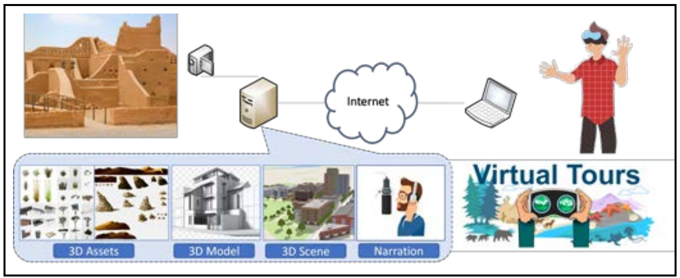


Fig. 2. Conceptual Model of How Virtual Tourism Concept is Created and Experienced.

A. *How VT and DH Works?*

Before the user can have a remote experience, the virtual environment must be created using CGI (Computer Graphics Interface). Fig. 2 shows a conceptual model of how the content is created for a virtual experience. For the non-immersive experience, a simple or 3D camera can be used to capture images and composed them together to create a visual tour using available toolkits and applications. 2D/3D assets are created using graphics tools, transformed into a 3D scene and then a virtual world is created. After that, narration is often added to guide the user.

For semi-immersive experiences, 360-degree cameras can be used which provide a panoramic view of the actual place. A user can connect to this 360-degree camera using the Internet and view it using a head-mounted display with motion sensors. When a user moves the sensors observe the movement and send to the controller, which calculates the viewing angle and direction and changes the display image accordingly. YouTube360 and Metaverse by Facebook are examples of these experiences.

Developers also use extensive CGI to create a 3D model of a place, such as, a palace or a museum, and augment it with the real images captured in real-time [35]. They use both audio and video streams to provide a fully immersive experience. The view in the 3D model is changed according to the movement of the user, which is recorded using gyroscope, magnetometer and accelerometer present in the head-mounted display or the mobile phone. Stereo surround-sound is used to provide directional sounds to the user with the built-in speaker in the head-mounted display. Similarly, for interaction with the environment, users can use a magnetic, ultrasonic, or gesture-based tracking systems to grab, and/or displace objects, actuate menus and/or perform other activities. The system receives the input signals, processes the interaction model, and displays the activities to the user at runtime. Instructions and detailed information about events, places, milestones, and important sites are also augmented in the 2D/3D world for guiding the users about the history or the significance of the place.

B. VR Equipment

1) Display devices: The quality of immersive in VR is defined by the display devices. A common monitor screen can provide a virtual experience; however, it would be a non-immersive impression [45]. A mobile phone can provide a semi-immersive experience, as it has motion sensors, which can detect changes in orientation and direction, and may respond by changing the view for the user [46]. The display devices use split screens to emulate a wide field of 360-degree stereoscopic view. This gives the user an illusion of looking around in the environment. Virtual Box [44] and Google Cardboard [43] are examples of such a device; however, they only provide a casing for the mobile device and help the user mount on the display device [44].

For a fully immersive experience, a range of head-mounted display devices are available, that contain sensors, cameras, location devices and controllers, built-in inside the device. Table I shows the comparisons of some of the famous head-mounted display devices, which are commonly known as VR kits.

As shown in Table I, Google Cardboard, and VRBox and similar devices are very cheap as they provide a method of holding a mobile phone for proper viewing the virtual environment. An elastic band is used to mount the device over user's head, or the user has to hold it using hands which makes it ergonomically infeasible. Some Bluetooth based controllers can be used to provide input; however, the experience in semi-immersive. On the other hand, HTC Vive Pro 2, HP Reverb G2, and Pimax Vision 8K X VR Headsets have ergonomic designs and are very comfortable with adjustable straps, which makes them very expensive. The other factors that make some of these devices expensive, is the display resolution and refresh rate. Pimax Vision 8K X VR Headset has the best display quality, however, high-quality CGI means an expensive GPU is also required. Every headset (in Table I) that is providing a fully immersive experience, requires a fast computer with a high-end GPU device except Oculus Quest 2, which can work as a stand-alone. It has its own processor and memory and can work with or without a computer. Oculus Quest 2 does not require extra devices, such as base stations, to track the headset and/or the controllers and provides flexible mobility using wireless communication. Oculus Rift, HTC Vive Pro 2, HP Reverb G2, HP Windows Mixed Reality, and Pimax Vision 8K X VR Headsets are connected through a wire to the computer; therefore, mobility is limited. However, wireless connectivity is being made available by various vendors to compete with the acceptability of Oculus Quest 2.

TABLE I. A COMPARISON OF VARIOUS HEAD-MOUNTED DISPLAY FOR VR

S. No	VR kit	Characteristics	Experience
1	Oculus Rift [36]	• Resolution 1280x1440, 80 Hz • No computation or memory resources • Accelerometer, proximity, magnetometer, gyroscope • Guardian system for headset and controllers tracking • Buttons and gestures-based input • Stereo speakers, positional audio • No Battery • Inexpensive	Fully immersive
2	Oculus Quest 2 [37]	• Stand-alone, • Resolution 1832x1920 per eye, 80 Hz • Qualcomm® Snapdragon™ XR2 processor with 128/256 GB ROM • Accelerometer, proximity, magnetometer, gyroscope • Guardian system for headset and controllers tracking, camera-based outside tracking, GPS • Buttons and gestures-based input • Stereo speakers, positional audio • 2–3-hour Battery life • Inexpensive	Fully immersive
3	HTC Vive Pro [38]	• Resolution: 2448x2448 per eye, 90/120 Hz, 5K display • 100-degree field of view • No computation or memory resources • G-sensor, proximity, gyroscope, IPD sensor • Guardian system for headset and controllers tracking, infrared camera-based outside tracking • Buttons and gestures-based input • Stereo speakers, positional audio • No Battery • Very Expensive	Fully immersive
4	HP Reverb G2 [39]	• Resolution: 2160 x 2160 per eye, 90 Hz, 2 LCDs • No computation or memory resources • G-sensor, proximity, gyroscope, IPD sensor • Guardian system for headset and controllers tracking and camera-based tracking • Stereo speakers, positional audio • No Battery • Expensive	Fully immersive
5	Sony Playstation VR [40]	• Resolution: 960x1080 per eye, 90/120 Hz • Graphics Controller Box for screen mirroring and sound processing • Accelerometer, gyroscope • Positional tracking with 9 LEDs via PlayStation Camera	Fully immersive

		• Stereo speakers, positional audio • No Battery • Expensive	
6	HP Windows Mixed Reality Headset	• Resolution: 1440 x 1440 per eye, 90 Hz, 2 LCDs • No computation or memory resources • Windows Mixed Reality inside/out 6 DOF motion tracking, gyroscope, accelerometer, and magnetometer • Front-facing camera tracking • Stereo speakers, positional audio • No Battery • Expensive	Fully immersive
7	Pimax Vision 8K X VR Headset [41]	• Resolution: 3840 x 2160 per eye, 60-114 Hz (2560x1440), 170-degree FOV • No computation or memory resources • 9-axis accelerometer • SteamVR tracking • Buttons and gestures-based input • Stereo speakers, positional audio • No Battery • Very Expensive	Fully immersive
8	DPVR Headset [42]	• Resolution: 1280x1400 per eye, 72 Hz, 110-degree FOV • No computation or memory resources • 3 DoF Non-positional tracking • SteamVR platform • No Battery • Very Cheap	Semi immersive
9	Google Cardboard [43]	• Uses mobile phone as a display • No computation or memory resources, • Very cheap	Semi immersive
10	VR Box [44] and similar	• Use mobile phone as a display • No computation resources, No memory, • Bluetooth based input devices are available • Inexpensive	Semi immersive

2) Sensors: A variety of sensor technologies are needed to provide input to VR/AR system. These sensors include object and motion tracking, directional sensing, visual sensing, and audio interfacing [46]. The technology is ever evolving by the introduction of new types of sensors and different types of haptics for touch, smell, and heat-based feedbacks. A conventional VR/AR system uses G-sensors, such as, accelerometer, magnetometer, and gyroscope to detect inertial movement and direction. Indoor and outdoor GPS systems are used to identify the location of the user, directional mics are used to understand what the user is saying or hearing and from where, and cameras are used to identify what the user is seeing. State-of-the-art devices are using more complex sensors, such as, time-of-flight sensors, heat mapping, structured light sensors, etc., which are product of integrated basic sensors mentioned above. Table II provides the details of some common sensors and tracking systems used in AR/VR technologies.

TABLE II. VARIOUS SENSOR USED IN AR/VR SYSTEMS AND THEIR USAGE

S. No	Sensor	Usage
1	Accelerometer	Movement tracking in X, Y and Z dimensions
2	Gyroscope	Sense angular velocity or rotation motion
3	G-sensor	Force of movement in gravitational units
4	Magnetometer	Direction tracking
5	Proximity	Measure distance of object in front
6	Light sensor	Measure the luminescence
7	IR sensor	Senses the invisible light (infrared) for proximity and motion detection
8	Depth sensors	Camera based sensor to detect the depth in the recorded image to identify which object is near and which is far away
9	Eye-tracking	Track the movement of eyes to identify where the user is looking
10	Directional microphone	A pair of mics which can identify where the sound is coming from using Doppler effect.
11	Inertial Measurement Sensor	An integrated accelerometer, gyroscope and magnetometer sensor
12	Time-of-Flight sensor	Laser or IR based distance detection for finding range between object and the camera. Used for robot navigation, vehicle monitoring, people counting, and object detection
13	Object & Gesture Tracking	Camera and image processing-based system used to track objects and user for recognition movement and gestures
14	Ultra-sound sensors	Ultrasound sensors are mainly used to detect proximity and distance
15	Thermal sensor	Helps in detecting heat signatures to differentiate between users and various objects.
16	Ambient light sensor	Helps in detecting heat signatures to differentiate between users and various objects.
17	GPS	Global Positioning System uses satellite-based location calculation and only works in open environments
18	Indoor GPS	To enable indoor location, Bluetooth or Wi-Fi beacon-based solutions are used where a device uses a triangulation method to calculate its own location

3) Cameras: The VR world is artificially created using CGI which can provide a 2D view for non-immersive display or a 3D view, which can be used to provide a fully immersive experience. But to create a whole 3D model, a lot of effort is required, which may not be realistic. To make the world more realistic, it is augmented by mixing the real-world object and scenes. Cameras are used to capture these images and then the virtual world is augmented with these captured images. Hence, the quality of the camera has significance in creating a realistic environment along with the 3D modelling and rendering process.

A simple camera can be used to take pictures of an environment from different angles and the images can be stitched together to provide a 360-degree view, which is time-consuming. These days, cameras are available that take a 360-degree picture of an environment and output a 3D world to be viewed at runtime or to be saved to create realistic mixed realities. A 3D camera is either omnidirectional or composed of many small cameras. It captures multiple images at the same time from different angles. The software, either inside the camera or on a computer or smartphone, orchestrates a spherical image by fusing the images. VR headset can be used to view these spherical images as it allows the user to move inside the image and feel an immersive experience. A comparison of some famous 360-degree cameras is provided in Table III.

TABLE III. COMPARISON OF FAMOUS 360-DEGREE CAMERAS

Q. No	Camera	Features
1	Insta360 One X2	• 2x 5.7K lenses • 6080x3040 360-degree recording • 6-axis gyroscope • 360 directional focus audio • Bluetooth, Wi-Fi, USB connectivity • MicroSD card storage • Less Expensive
2	GoPro Max	• 1x 5.6K lens • 16.6mp 360-degree recording • GPS • 4 channel microphones audio • Bluetooth, Wi-Fi, USB connectivity • MicroSD card storage • Less Expensive
3	Ricoh Theta Z1	• 2x 4K lenses • 23mp 360-degree recording • GPS • 6x microphones audio • Bluetooth, Wi-Fi (web server), USB connectivity • 50 GB internal storage • Expensive
4	Samsung Gear 360	• 2x 15mp lenses • 3840x1920 360-degree recording • Accelerometer, Gyroscope • Bluetooth, Wi-Fi, USB, NFC connectivity • Compatible with High-end Samsung Mobile phones • 1 GB internal storage with 128 GB using MicroSD • Cheap
5	Vuze XR	• 2 x Sony 12MP IMX-378 fisheye lenses • 3840x1920 360-degree recording through Ambarella H2 video processor • Accelerometer, Gyroscope • Wi-Fi, USB connectivity • 4x microphones • Removeable MicroSD card • Cheap
6	Nokia OZO	• 8 2Kx2K ISO 190-degree lenses • 9 Capture up to 12K30 x 12K30 Video/Stills • 500 GB SSD Module for Recording • Automatic stitching • HDMI output • Omnidirectional microphones • Extremely Expensive
7	Gopro Odyssey	• 16 x 2.7K Lenses • MicroSD card storage • Automatic stitching • USB cable for connectivity • 16-mono microphones • Expensive
8	Kandao Obsidian GO 360° 3D VR Camera	• 8 x 6K f/2.8 195° Fisheye Lenses • Capture up to 12K30 x 12K30 Video/Stills • Internal 8TB SSD Module for Recording • 8 x 6K f/2.8 195° Fisheye Lenses • Automatic stitching • Wi-Fi 6, Bluetooth 5, Gigabit Ethernet • 4-directional microphones • Extremely Expensive
9	Panono 360° 108MP Camera	• 6 x 3MP Cameras with 360° 108MP Still Image recording • Image Requires Stitching • Wi-Fi Connectivity • Expensive
10	Z CAM V1 Spherical VR 360 Camera	• 10 fisheye lenses with 190-degree view, each • 7K 360-degree video recording • Automatic stitching • Live video streaming • 4 built-in microphones • Very Expensive
11	Z CAM V1 Pro Cinematic VR Camera	• 9 MFT Lenses with 190-degree view • Automatic stitching • Live video streaming • 4-directional microphones • Extremely Expensive

4) Input devices: As with any system, user interaction is mainly based on the input devices being used. Although in VR most of the input is delivered through sensors; there are many input devices and controllers available for the users. Most of these input devices are only compatible with specific head-mounted displays, but some generic input devices are also available. Table IV presents some of the input devices that are being used by VR users and their classification based on the sensor technology they are equipped with.

TABLE IV. THE FAMOUS INPUT DEVICES OR CONTROLLERS FOR VR/AR KITS

S. No	Input Device	Class	Features
1	Oculus Controllers	Position and Orientation Tracking	IMU sensors, IR-Led- based tracking ring (for fingers) and camera-based hand tracking, HD haptic feedback, buttons and thumbstick
2	HTC Vive Controllers	Position and Orientation Tracking	24 IMU sensors, buttons, multi-function trackpad, dual-stage trigger, IR-based hand tracking, HD haptic feedback and a rechargeable battery
3	HP Reverb Controllers [39]	Position and Orientation Tracking	IMU sensors and active LEDs based tracking and IR camera-based hand tracking, haptic feedback, buttons and thumbstick
4	PlayStation VR Aim Controller [40]	Dual Camera-based Object Tracking	2x cameras, Hand tracking, gesture recognition
5	HP Windows MR Controllers	Position and Orientation Tracking	IMU sensors, multi-function trackpad and IR-led and camera-based hand tracking, HD haptic feedback, buttons and thumbstick
6	Polhemus Fastrak	Magnetic Tracker	3D digitizer and a quad receiver motion tracker, which computes the position and orientation of a small receiver
7	Mattel Power Glove	Acoustic (Ultrasonic) Trackers	Buttons, detects yaw, pitch and roll of hand, uses fiberoptic sensors to detect finger flexure to 256 positions per finger for four fingers. Uses sonics to calculate X, Y and Z location also

8	NAC Eye Mark Eye Tracker	Eye Tracking	Uses Pupil / Cornea Reflection to detect eyeball movement. Provides video footage of the field of view of the user in real-time
9	VPL Data Glove	3D input device: Glove	Camera and image processing-based tracking of the specific-colored glove to detect gestures as input.
10	Virtex Cyber Glove	3D input device: Glove	Uses electronic joint-angle sensors, electro-magnetic position tracker, and virtual hand control software detect gestures as input
11	Joystick based upon 6DOF	3D input device: mouse	Uses force-sensors to detect change in momentum and identify movement in all six directions.
12	Dexterous Hand Master (DHM)	3D input device: Dexterous manipulators	Extremely accurate movement and gesture-detection device used for medical and scientific training. Based on IMU sensors, position trackers and/or visual sensors.

C. Toolkits and Libraries

The tools and technologies are the building block of the virtual experience, and they are evolving with new features every day. They allow the user to have an immersive experience of the virtual or real world and provide a realistic interaction with the environment by constantly sending and receiving data from the devices to the multimedia system and vice versa. Another very important aspect of VR/AR systems is the 3D design and rendering model inside the server or mobile phone that enables this real-time interaction. These 3D realistic models and systems are created by graphic designers and software engineers using various programming frameworks.

To prompt development of 3D object and scene, various application and tools are available for the developers. For modelling object CGI tools, such as, 3D Studio Max, Maya, and Adobe Illustrator, are used. For building the scene and implementing the concepts of physics using complex mathematics of collision meshes, boundary detection, object interaction, differential equations, and fluid dynamics, the need for a support library is eminent. For prompt development, developer make use of 3D graphics engines, such as, Unreal Engine, Unity, Amazon Lumberyard, Unigine, if Tech 5, 3ds Max Design, ApertusVR, etc. These engines provide a VR SDK, which allows them to design, build, and test their VR environments.

Similarly, some toolkits are also available, which provide a quicker way to create VR games and environments [47]. They provide a collection of useful scripts, 3D assets and templates, interaction models, and libraries to interface various sensors, displays and actuators in a VR application.

VRSciT project provides a toolkit to explore new approaches in educational tourism using a VR environment [48]. TIDE Toolkit is a tourism toolkit for European Maritime and Underwater Cultural Heritage [49]. Digital Trail is another toolkit that allows developers to build their VR/AR content as an App and share the App with the users [50]. VITAKI is a toolkit designed to create VR applications and games with specific libraries for interaction with vibration and haptics [51]. VR Mini-Degree is another toolkit that provides tools to build VR games and applications using Unity. Samples of first shooter games, 360-degree space view and other examples are available to reduce the learning time for the developers. The tourism department in Northern Ireland has published a Cultural heritage toolkit for developing Irish cultural heritage experience for the tourists [77]. These are set of guidelines for developing virtual tourism systems. A similar toolkit is also published by Scotland [53] and a set of guidelines are published in [52] for promoting sports tourism.

For developing VR/AR systems, many developers are using VRTK, which is a VR toolkit for Unity Engine [47]. VRTK library consists of numbers of solution or mechanisms, like movement in VR, interaction with object like touching and catching, and 2D and 3D control like button, lever, and another objects. There are various libraries that provide ease in implementing tracking in AR/VR systems, such as, Qualcomm’s Vuforia and ARToolKit, and the BuildAR authoring toolkits.

COLIBRI VR is an open-source VR toolkit for capturing, modelling, and rendering real-world scenes in a VR environment [54]. For visualizing data in an immersive environment, authors in [55] have developed a VR toolkit for building and analyzing datasets and processed results. OVR toolkit provides a desktop view in a VR experience by augmenting a VR environment with real-life desktop objects [56]. VR performance toolkit [57] provides models and algorithms to improve performance in VR-based games. POV’s VR toolkit provides 3D assets and pre-build environments to create VR Apps in a matter of hours. Vrui VR toolkit provides functionalities, such as, portability, scalability, and fast development for VR applications. XR Interaction toolkit provides a cross platform, input, haptics, interaction, and feedback procedures for the VR environment. BlocklyXR is another toolkit which provides a visual programming environment for implementing realistic interactions in developing digital storytelling VR applications [58].

There are many other toolkits available for developers for visualizing data, developing educational applications and training exercises [59]. With time, better and powerful tools and technologies will be available for the developers to create highly immersive VR/AR/MR experiences.

IV. Survey Analysis

To identify the awareness of the field of virtual tourism and digital heritage among the users of the Internet, we opted to conduct a survey analysis. For better response, we limited our survey to less than ten (10) questions. The questions asked in the survey were all close-ended questions with mandatory responses. Table V shows the questions asked in the survey. We also identified the region from where the response was collected using the IP-based location services. We divided the responses according to the following regions: (1) Far east and south-east Asia (including Australia), (2) Central & south Asia, (3) Middle east & northern Africa, (4) Europe, (5) North America, and (6) Latin & south America. The survey was distributed through academic links; hence, most of the responses are from people who are associated with the education sector. We receive almost 550 responses, out of which 534 are valid responses.

TABLE V. SURVEY QUESTIONS AND THEIR POSSIBLE RESPONSES

Q. No.	Questions	Possible Responses
1	Gender	Female, Male, Not Specify
2	Age Group	Baby Boomers, Gen X, Millennial, Gen Z
3	Are you aware of Vitual Tourism?	Yes/No
4	Are you aware of Digital Heritage?	Yes/No
5	Have you ever experienced Vitual Tourism?	Yes/No
6	Have you ever experienced Digital Heritage?	Yes/No
7	Are you willing to experience Vitual Tourism?	Yes/No/May be
8	Are you willing to experience Digital Heritage?	Yes/No/May be

A. Demographics of the Responses

Demographics of the collected data is shown in Fig. 3. Out of the 534 responses, 21% identified themselves as male, 10% as female, while 69% preferred not to disclose their gender (shown in Fig. 3(a)).

For the age groups, 2% only identified themselves as Baby Boomers (above 56 years old), 17% as Generation X (43-55 years old), 44% as millennials (aged 30-42 years), and 37% as Generation Z (20-29 years old).

Region-wise, we got 16% responses from Far East and south-east Asia, 27% from central & south Asia, 21% from middle east & northern Africa, 16% from Europe, 13% from north America, and 7% from Latin & south America, as depicted in Fig. 3(b). The demographic results show that although the responses are not a lot, but the sample covers a major portion of the world population with people from all age groups.

B. Awareness of VR and DH

The awareness of existing applications for virtual tourism and digital heritage is shown in Fig. 4(a) and 4(b) respectively. In the overall scenario, more people are aware of virtual tourism than digital heritage. But comparatively, Millennials and Generation Z have more awareness than the older age groups. It happens because of the technological savvy nature of the younger generations.

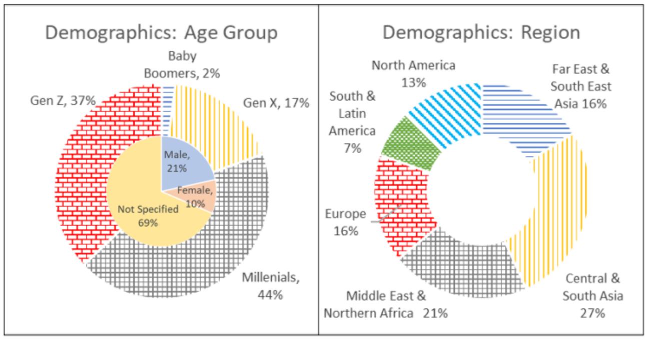


Fig. 3. Demographics of the Responses for the Survey. (a) Age Groups and Gender (b) Regions.

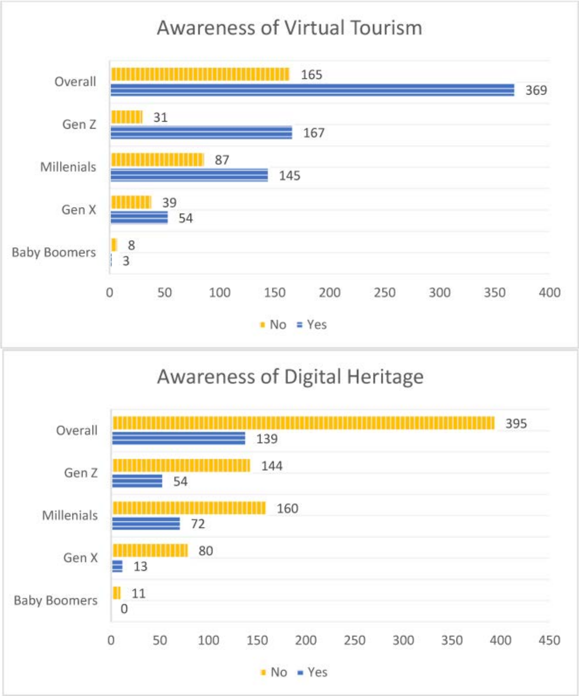


Fig. 4. Response Statistics about the Awareness of (a) Virtual Tourism and (b) Digital Heritage.

C. Experience with VR and DH

Fig. 5(a) and 5(b) show the results of the people who have actually experienced the virtual tourism and digital heritage applications, respectively. We see similar trends, as in Fig. 4. But if we compare the results of Fig. 4 and Fig. 5, we see that although many people are aware of the virtual tourism; however, fewer have experienced it.

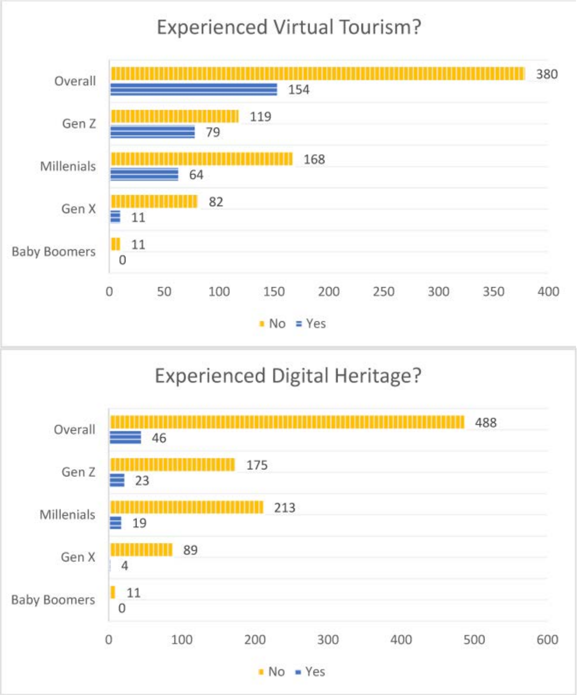


Fig. 5. Response Statistics about the How many People have Experienced of (a) Virtual Tourism and (b) Digital Heritage.

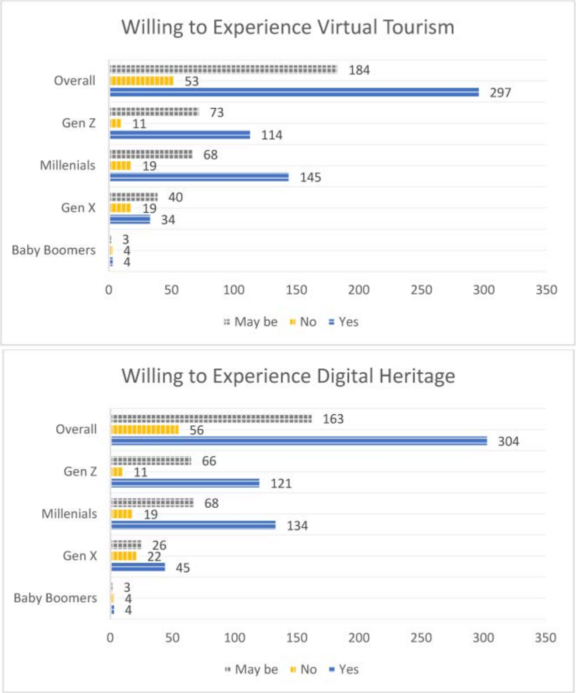


Fig. 6. Response Statistics about the How Many People have shown Willingness to Experience (a) Virtual Tourism and (b) Digital Heritage.

In case of digital heritage, the values are even lower. Again, the younger age group is more interested in the concept of virtual tourism as compared to the older age groups.

D. *Willingness to Experience VR and DH*

Finally, Fig. 6 shows the willingness of people to experience the applications of virtual tourism and digital heritage. It can be observed that most people have shown interest in the field, and they plan to visit the website and use applications that provide them with a virtual experience. However, many people still prefer the real traditional experience. The reasons that motivate people to experience the virtual environment or even the real environment with a remote axis are discussed in the next section. We also discuss the demerits of virtual tourism and digital heritage.

V. Pros and Cons of VR & DH

Although, there are many advantages of virtual tourism and digital heritage; however, the experience lacks certain supplementary aspects that have real importance and impact on the tourists. Some of these issues are discussed in the following subsections as advantages and disadvantages.

A. *Advantages*

1) Safety and security: One of the biggest advantages for the person who chooses virtual tourism over physical one is the safety and security of the individuals. Many tourist destinations are filled with cheaters and scammers and even kidnappers. There have been many incidents where people were abused mentally and physically [60]. VT provides a safe and secure experience from the protected home environment.

2) Control on view: In a VT, the users may have control over what they want to view and what they want to avoid. For example, some people are introverted and lack skills in interacting with other people and consider them as aliens, which helps them to avoid self-embarrassing behavior. VT allows the users to eliminate these experiences and focus on what they want to view and how they want to view the remote environment.

3) Cost effectiveness: Although tourism is one of the financially stable industries; however, not everyone is blessed with enough fortune to visit faraway places due to the cost of traveling, accommodation, and fares. VT and DH provide a feasible opportunity to the less privileged people as the virtual tourism cost considerably less than the traditional tourism. Though, the experience is not similar, but the users can save on cost by opting for VT. Some of the VTs are available free of cost for promotional purposes, while digital heritage sites are supported by governmental funding to make them accessible for free of cost.

4) Less impact on vulnerable destinations: Natures beauty is significantly destroyed whenever a location becomes a tourist spot. Tourism industry requires accessibility, accommodation, and utilities for the visitors. Hence, roads are made, hotels are erected and the people who visit later may damage the environment. In contrast, virtual tour can be recorded without endangering the environment and natural beauty can be preserved.

5) Preserve environment: The direct benefit of DH is the preservation of the cultural, historical, scientific, educational, and linguistic resources. For example, many artifacts lose their originality when tourists visit the historical sites. Likewise, access to ancient sculpture, sites, and books can destroy important parts and evidence of the history. Therefore, access to these resources is enabled via virtual, augmented and/or mixed reality.

6) Try before buy: VT provides the opportunity for tourists to first try the cheaper virtual experience before deciding to travel to the actual destination. This avoids or at least limits bad experiences for the travelers. They can try the virtual tour by experiencing it and make an educated decision to save time and money.

7) Permanent memory in high resolution: Everyone likes to take pictures and make videos of the place they visit. It helps them to recall their memorable moments and to capture the beauty and details of the visited places. VT and DH allow them to have high-quality pictures and videos of their experience. Furthermore, they can experience the environment again, anytime they want.

8) Visit the past, the present and even the future: With the Advent of virtual tours and digital heritage, a person is able to walk among dinosaurs, witness the landing on the moon and even walk on water. A user can experience the historical achievements of mankind and places which have been destroyed by time and urbanization. Similarly, a person can revisit his experience of the past, which s/he enjoyed with the loved one. Envisaging the human imagination, we can even visit the future world, which we want to create for our next generations.

B. Disadvantages

1) Lack of physical interaction: In real tourism, people share cultural knowledge and experience with other people through physical interaction by participating in the activities. When tourists visit any place, they take part in the cultural events, chat with the local people and gain valuable insights from their daily life. Such experience becomes long lasting, and they can share with other people effectively. Visiting and experiencing various diverse cultures improves humans' inter-personal skills. However, in a virtual experience, a user is not able to experience the diverse culture of the world and remain unknown of the fellow human beings.

2) Artificially enhanced scenery: Some of the virtual tourism application provide CGI (Computer-generated imagery) enhanced scenery, which is quite different from the actual environment or destination. Mostly, this is done to motivate the user for marketing purposes; however, this may come as fraud or intentionally mislead the users from reality. Therefore, some of the users may not opt for virtual tourism and prefer to physically visit the destination. In case, if a user witnesses such a difference experience, then user will be reluctant to experience virtual tours in the future.

3) Lack of relaxation: Some people travel for sight-seeing, as well as relaxation; however, the virtual tourism may not provide relaxation. On the other hand, some people feel tired after using the AR/VR systems due to the artificial visual movements. People who are claustrophobic or suffer from vertigo and epilepsy, would not be able to have a decent experience from the virtual tours.

4) Anti-globle village concept: The world is fast becoming a global village with technological advances, such as, fast travelling, Internet, social media, etc. People are becoming close to each other as they interact through these platforms. However, travelling is still the most significant way that makes people be a part of that global village. In virtual tourism, physical contact is missing, which enable the sense of being together. Younger generations are already distancing themselves from actual physical contacts and rely on online communication and the concept of VT will only augment this problem.

5) Change of environment: One of the most common reasons for travelling is a change in environment. Mostly, people have a busy schedule with consistent daily timetable, which they follow for months and years. They start feeling bored and wearied by this perpetual routine. Hence, they look for vacation with a change in scenery and itinerary in mind. Virtual tourism may create a change of experience for a few hours, but the minute change does not impact the person psychologically and they do not feel thrilled and animated.

6) Lack of availability to everyone: The technologies which are used to enable VT and DH are still expensive and their application is still not available for everyone to explore. Although most of the applications are free, but the cost of Internet and display devices/mobile phones (with high-end computer graphics) affects the availability of virtual tours.

7) Lack of financial gains: A tourist destination can earn a lot due to the visiting travelers through tickets, most of which is used for the management of the destination. At the same time, there are some supplementary industries that observe an increase in business due to the increase in travelling such as hotels, restaurants, travelling agencies, airlines and bus service, guides, etc. The concept of VT can reduce the numbers of tourists, which could have a severe impact on the earning of the tourist destination, as well as the supplementary industries.

VI. Related Work

In this section, we present similar works done by various researchers in the context of virtual tourism and digital heritage. In literature, we found different surveys and review papers [9-10] [27] [32-33] [61] [64-72] [74] discussing the feasibility, business value, social impact, marketing trends, technological evaluation, challenges, and issues. Other research articles [62] [63] [34] [73] [75-76] proposed diverse solutions pertaining to virtual, augmented, and mixed realities, while targeting virtual tourism and/or digital heritage. A brief statistical overview of existing studies in terms of citations and year of publication is provided in Table VI, which explore the surveys and reviews published in the field of virtual tourism and digital heritage. It is evident that most of the related works targeting virtual tourism as compared to digital heritage. Moreover, many people are citing the work done in VT, which shows lack of efforts in DH domain.

Verma et al. [10] has provided a comprehensive review of the literature related to virtual tourism with focus on the technologies, methodologies, and impact of VR/AR on tourism industry. They have also identified potential stakeholders and presented a quantitative and qualitative analysis of published research in tourism domain. Fan et al. [32] has analyzed the impact of immersive technologies on the virtual tourism by examining the social interactions and experience feedbacks from the users. Liang et al. in [61] reviewed augmented reality for tourism using cluster-based analysis with four empirical studies. They have based their recommendations on identified future directions to propose further investigation in the emerging area of gamification and explored the potential negative consequences of augmented reality. In [62], authors have explored the alternative ways of travelling during the pandemic and identified virtual tourism as a solution. The author of [9] has reviewed around 32 journal papers, published during the years 2000-2018, to not only identify the impact of VR and AR on virtual tourism but also suggested future directions. Similarly, Loureiro et al. [63] have used citation network analysis and text-mining to study 56 journal and 325 conference papers for identifying the requirements, challenges, solutions, and future trends in the field of virtual tourism using AR/VR technologies. Egger et al. in [64] have provided a literature review on AR/VR technologies and identified the strengths and weaknesses in the context of accessibility and marketing towards tourism.

TABLE VI. LITERATURE REVIEW OF THE VIRTUAL TOURISM AND DIGITAL HERITAGE REVIEW AND SURVEY PAPERS

Authors & Reference	Citations	Publication Year	Field
Verma et al. [10]	0	2022	VT
Fan et al. [32]	2	2022	VT
Liang et al. [61]	19	2021	VT
Riesa et al. [62]	5	2020	VT
Çeltek [9]	2	2020	VT
Loureiro et al. [63]	225	2020	VT
Egger et al. [64]	6	2020	VT
Pestek and Sarvan [69]	33	2020	VT
Loureiro and Correia [70]	4	2020	VT
Kononova et al. [71]	4	2020	VT
Beck et al. [65]	129	2019	VT
Yung and Khoo-Lattimore [66]	435	2019	VT
Wei. [67]	88	2019	VT
Claire and David [68]	84	2003	VT
Batchelor et al. [72]	4	2021	DH
Champion and Rahaman . [27]	21	2020	DH
Abdelhamid and Galal [73]	5	2019	DH
Bekele et al. [33]	411	2018	DH
Münster S. [74]	21	2017	DH
Münster et al. [34]	20	2016	DH
Tammaro and Maria [75]	10	2016	DH
Loannides et al. [76]	39	2014	DH

Beck et al. [65] have performed a comprehensive review on VR for tourism and discovered that the VR is complementing tourism as a marketing tool rather replacing it. However, this paper was published before the COVID pandemic and since then the trends have changed as suggested in [10]. Yung and Khoo-Lattimore, in their review paper [66], have identified various factors, methodologies, and class of virtual technologies that can benefit VT. They have also identified research gaps and potential tourism sectors in this field. A critical review is presented in [67] to understand the research progress on VR/AR in tourism and hospitality from 2000 to 2018. Based on the review of 60 journal articles, they have recommended research directions in VR/AR with their management. Authors of [68] have reviewed the existing and potential exploitation of Virtual Learning Environments (VLEs) within hospitality, leisure, sport, and tourism. Other research studies in [69], [70], and [71] provided short reviews on identifying enabling technologies of VR and AR in transition from tourism to e-tourism.

Bekele et al. in [33] have reviewed the trends in VR/AR/MR that can benefit the cultural heritage paradigm. They also identified application areas in digital cultural heritage and identified appropriate technologies for various digitizing cases in preserving heritage and culture. Batchelor et al. in [72] have presented a literature review on digital and smart heritage by identifying the requirements, challenges, and analyzing the complementary technologies. Champion and Rahaman in [27] provided a collection of websites and repositories on 3D model of heritage sites. They also reviewed various platforms that support the digital heritage and provided guidelines for future proofing and preserving the heritage using digitization. Authors in [73] have discussed digital freehand sketching, digital measurements, photographic techniques for creating panoramas, 3D models and interactive tours, interactive virtual tours, VR techniques and 2D and 3D model creation for legacy digitization and other trends. Similarly, researchers have provided various literature reviews that discuss the topics and research domains [74] and identified challenges and solutions for preserving digital heritage [34], [75], [76].

Review papers to analyze the AR/VR technologies in the field of virtual tourism are many; however, the technologies are changing with the advancements in sensing, tracking, and computing capabilities. We noticed a lack of interest and limited research work in digital heritage domain compared with virtual tourism. As digital heritage and environment for virtual tourism have similarity in media creation, interaction, and immersive experience; hence, a survey to review the AR/VR technologies and application for both VT and DH was missing. This paper has fulfilled this gap and provided a single knowledge source for people working in the blended domain.

VII. CONCLUSION

In this paper, we have reviewed and analyzed the concepts of virtual and augmented realities, which have a significant role in the field of virtual tourism and digital heritage. We have presented an in-depth analysis about the commercial technologies and web/mobile application, which can benefit the research community, especially, novice researchers in comprehending the current spectrum in the field of virtual tourism and digital heritage. We have presented the survey analysis of the awareness, experience, and willingness of using virtual tourism and digital heritage by the users of various age groups. The results have shown that the young generations are excited to experience these technologies. In the end, we also provide the advantages and disadvantages associated with virtual and remote tourism.

ACKNOWLEDGMENT

The work presented in this paper is funded by the Deputyship for Research and Innovation, Ministry of Education, Kingdom of Saudi Arabia, under the project No (20/16), titled: “Virtual journey into the history of the Kingdom of Saudi Arabia”. We would also like to extend special thanks to all the people who participated in the conducted survey.

REFERENCES

[1] Farshid, Mana, Jeannette Paschen, Theresa Eriksson, and Jan Kietzmann. "Go boldly!: Explore augmented reality (AR), virtual reality (VR), and mixed reality (MR) for business." Business Horizons 61, no. 5 (2018): 657-663.

[2] Alcantud, Francisco, Gerardo Herrera, Gabriel Labajo, I. Dolz, C. Gayá, V. Avila, A. Blanquer, Jose Luis Cuesta, and J. Arnáiz. "Assessing virtual reality as a tool for support imagination." In International Conference on Computers for Handicapped Persons, pp. 143-144. Springer, Berlin, Heidelberg, 2002.

[3] Jung, Timothy, and M. Cluaudia tom Dieck. "Augmented reality and virtual reality." Ujedinjeno Kraljevstvo: Springer International Publishing AG (2018).

[4] Narayan, Paresh Kumar, Dinh Hoang Bach Phan, and Guangqiang Liu. "COVID-19 lockdowns, stimulus packages, travel bans, and stock returns." Finance research letters 38 (2021): 101732.

[5] Flix Ritcher, "Pandemic Could Set Tourism Sector Back by $1 Trillion", Statista, https://www.statista.com/chart/22689/global-international-tourism-receipts/.

[6] "Global economy could lose over $4 trillion due to COVID-19 impact on tourism", COVID-19 and tourism - An update, Assessing the economic consequences, UNCTAD/PRESS/PR/2021/023, 30 June 2021.

[7] Parry, Ross. Recoding the museum: Digital heritage and the technologies of change. Routledge, 2007.

[8] Ventura, Sara, Eleonora Brivio, Giuseppe Riva, and Rosa M. Baños. "Immersive versus non-immersive experience: Exploring the feasibility of memory assessment through 360 technology." Frontiers in psychology 10 (2019): 2509.

[9] Çeltek, Evrim. "Progress and development of virtual reality and augmented reality technologies in tourism: A review of publications from 2000 to 2018." Handbook of research on smart technology applications in the tourism industry (2020): 1-23.

[10] Verma, Sanjeev, Lekha Warrier, Brajesh Bolia, and Shraddha Mehta. "Past, present, and future of virtual tourism-a literature review." International Journal of Information Management Data Insights 2, no. 2 (2022): 100085.

[11] Global Virtual Tourism Market Research Report, Industry Analysis & Forecast (2022-2027). https://www.marketdataforecast.com/market-reports/virtual-tourism-market.

[12] The 360-degree virtual museum tour of The Louvre, https://360stories.com/paris/place/louvre-museum.

[13] Dubai 360 virtual tours, https://www.360virtualtour.co/dubai-360-virtual-tours/.

[14] An interactive journey to the Great Barrier Reef, https://attenboroughsreef.com/.

[15] The Great Wall of China, Google Arts and Culture, https://artsandculture.google.com/story/igVxCi6iJJ6CrA.

[16] Virtual tours: Buckingham Palace, https://www.royal.uk/virtual-tours-buckingham-palace.

[17] Yosemite Virtual Tour, https://www.virtualyosemite.org/.

[18] View Colosseum in 360 virtual tour, https://www.touristtube.com/Things-to-do-in-Rome/Colosseum-360

[19] Statue of Liberty :: 360° VR Panorama, http://www.samrohn.com/360-panorama/statue-of-liberty-new-york/.

[20] The Holy Land Places in 360-Degree Virtual Reality Tour, https://www.p4panorama.com/gallery-item/the-holy-land/.

[21] Experience Mecca in VR, https://www.oculus.com/experiences/gear-vr/1125286047502859/.

[22] Brave the Himalayas in 'Everest VR: Journey to the Top of the World, https://www.oculus.com/blog/brave-the-himalayas-in-everest-vr-journey-to-the-top-of-the-world/.

[23] 360° VR Space Safari, https://orsted.com/en/explore/space-safari.

[24] The 35 Best Virtual Tours Online So You Can Travel From Home, https://worldwidehoneymoon.com/best-virtual-tours-online-travel-from-home/.

[25] De Lusenet, Yola. "Tending the garden or harvesting the fields: Digital preservation and the UNESCO charter on the preservation of the digital heritage." Library Trends 56, no. 1 (2007): 164-182.

[26] Exploring World Heritage from home with UNESCO, https://en.unesco.org/covid19/cultureresponse/exploring-world-heritage-from-home-with-unesco.

[27] Champion, Erik, and Hafizur Rahaman. "Survey of 3D digital heritage repositories and platforms." Virtual Archaeology Review 11, no. 23 (2020): 1-15.

[28] Global Digital Hertiage, Harnessing the latest technologies, https://globaldigitalheritage.org/.

[29] Zhang, Aishan. "The Narration of Art on Google Arts and Culture." Johns Hopkins University 1, no. 1 (2020): 21828.

[30] These 12 Famous Museums Offer Virtual Tours You Can Take on Your Couch, https://www.travelandleisure.com/attractions/museums-galleries/museums-with-virtual-tours.

[31] Take Virtual Tours of These Stunning Libraries, https://ilovelibraries.org/article/take-virtual-tours-these-stunning-libraries/.

[32] Fan, Xiaojun, Xinyu Jiang, and Nianqi Deng. "Immersive technology: A meta-analysis of augmented/virtual reality applications and their impact on tourism experience." Tourism Management 91 (2022): 104534.

[33] Bekele, Mafkereseb Kassahun, Roberto Pierdicca, Emanuele Frontoni, Eva Savina Malinverni, and James Gain. "A survey of augmented, virtual, and mixed reality for cultural heritage." Journal on Computing and Cultural Heritage (JOCCH) 11, no. 2 (2018): 1-36.

[34] Münster, Sander, Mieke Pfarr-Harfst, Piotr Kuroczyński, and Marinos Ioannides, eds. "3D research challenges in cultural heritage II: How to manage data and knowledge related to interpretative digital 3D reconstructions of cultural heritage." (2016).

[35] Kim, Won S. "Computer vision assisted virtual reality calibration." IEEE Transactions on Robotics and Automation 15, no. 3 (1999): 450-464.

[36] Ben Lang, "Oculus Rift S Revealed with Inside-out Tracking, Resolution Bump, & New Ergonomics", https://www.roadtovr.com/oculus-rift-s-specs-release-date-announcement-gdc-2019/.

[37] "Introducing Oculus Quest 2, the Next Generation of All-in-One VR | Oculus". developer.oculus.com.

[38] HTC Vive Pro 2 Headset, https://www.vive.com/us/product/vive-pro2/overview/.

[39] HP Reverb G2 Headset, https://www.hp.com/us-en/vr/reverb-g2-vr-headset.html.

[40] Sony Playstation VR, https://www.playstation.com/ar-sa/ps-vr/tech-specs/.

[41] Pimax Vision 8K X VR Headset, https://vr-compare.com/headset/pimaxvision8kx.

[42] DPVR Headset, Shanghai Lexiang Technology, https://www.dpvr.com/en/.

[43] Yoo, Soojeong, and Callum Parker. "Controller-less interaction methods for Google cardboard." In Proceedings of the 3rd ACM Symposium on Spatial User Interaction, pp. 127-127. 2015.

[44] Fröhlich, Thomas, Dmitry Alexandrovsky, Timo Stabbert, Tanja Döring, and Rainer Malaka. "Vrbox: A virtual reality augmented sandbox for immersive playfulness, creativity and exploration." In Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play, pp. 153-162. 2018.

[45] Cashen, Dan, and Emily Robb. "Augmented Reality Human-Machine Interface: Defining Future AR System Technology." In SID Symposium Digest of Technical Papers, pp. 23-28. 2015.

[46] Riendeau, P. "Augmented and Virtual Reality: The next big thing in marketing." Sensors for AR/VR. Pressbooks (2017).

[47] VRTK - Virtual Reality Toolkit, A productive VR Toolkit for rapidly building VR solutions in Unity3d, https://vrtoolkit.readme.io/.

[48] Lu, Weijun, Guanyi Ma, Qingtao Wan, Jinghua Li, Xiaolan Wang, Weizheng Fu, and Takashi Maruyama. "Virtual reference station-based computerized ionospheric tomography." GPS Solutions 25, no. 1 (2021): 1-12.

[49] Sawant, Nikita. "The tele-immersive data explorer(TIDE): a distributed architecture for tele-immersive scientific visualization." Master's thesis, University of Illinois at Chicago, 2000.

[50] Prokopenko, Olha, Valentyna Rusavska, Nelia Maliar, Alisa Tvelina, Nataliia Opanasiuk, and Halyna Aldankova. "Digital-toolkit for sports tourism promoting." International Journal of Advanced Research in Engineering and Technology (IJARET) 11, no. 5 (2020).

[51] Martínez, Jonatan, Arturo S. García, Miguel Oliver, José P. Molina, and Pascual González. "Vitaki: a vibrotactile prototyping toolkit for virtual reality and video games." International Journal of Human-Computer Interaction 30, no. 11 (2014): 855-871.

[52] Prokopenko, Olha, Valentyna Rusavska, Nelia Maliar, Alisa Tvelina, Nataliia Opanasiuk, and Halyna Aldankova. "Digital-toolkit for sports tourism promoting." International Journal of Advanced Research in Engineering and Technology (IJARET) 11, no. 5 (2020).

[53] Scotland Tourism Toolkit, https://www.ukinbound.org/wp-content/uploads/2020/07/toolkit-scottish-tourism-recovery-23.6.20.pdf.

[54] de Dinechin, Grégoire Dupont, and Alexis Paljic. "Demonstrating COLIBRI VR, an open-source toolkit to render real-world scenes in virtual reality." In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 844-845. IEEE, 2020.

[55] Sicat, Ronell, Jiabao Li, JunYoung Choi, Maxime Cordeil, Won-Ki Jeong, Benjamin Bach, and Hanspeter Pfister. "DXR: A toolkit for building immersive data visualizations." IEEE transactions on visualization and computer graphics 25, no. 1 (2018): 715-725.

[56] Kelly, Ryan. "The Universe of You: Using Remote VR to Improve Psychoeducation Through Spatial Presence, Attention Allocation, and Interaction." In Play Therapy and Telemental Health, pp. 229-239. Routledge, 2021.

[57] Chu, Jin Teck. "VR-XPR: a rapid prototyping toolkit and framework for developing high performance virtual reality applications." Master's thesis, Iowa State University, 2001.

[58] Jung, K., Nguyen, V.T. and Lee, J., 2021. Blocklyxr: An interactive extended reality toolkit for digital storytelling. Applied Sciences, 11(3), p.1073.

[59] Kaser, David, Kara Grijalva, and Meredith Thompson. Envisioning virtual reality: A toolkit for implementing VR in education. Lulu Press, Inc, 2019.

[60] Ayob, Norizawati Mohd, and Tarmiji Masron. "Issues of safety and security: new challenging to Malaysia tourism industry." In SHS Web of Conferences, vol. 12, p. 01083. EDP Sciences, 2014.

[61] Jingen Liang, Lena, and Statia Elliot. "A systematic review of augmented reality tourism research: What is now and what is next?." Tourism and Hospitality Research 21, no. 1 (2021): 15-30.

[62] Riesa, Rafidola Mareta, and Alfatah Haries. "Virtual tourism dalam literature review." Jurnal Pariwisata Bunda 1, no. 1 (2020): 1-6.

[63] Loureiro, Sandra Maria Correia, Joao Guerreiro, and Faizan Ali. "20 years of research on virtual reality and augmented reality in tourism context: A text-mining approach." Tourism management 77 (2020): 104028.

[64] Egger, Roman, and Larissa Neuburger. "Augmented, virtual, and mixed reality in tourism." Handbook of e-Tourism (2020): 1-25.

[65] Beck, Julia, Mattia Rainoldi, and Roman Egger. "Virtual reality in tourism: a state-of-the-art review." Tourism Review 74, no. 3 (2019): 586-612.

[66] Yung, Ryan, and Catheryn Khoo-Lattimore. "New realities: a systematic literature review on virtual reality and augmented reality in tourism research." Current issues in tourism 22, no. 17 (2019): 2056-2081.

[67] Wei, Wei. "Research progress on virtual reality (VR) and augmented reality (AR) in tourism and hospitality: A critical review of publications from 2000 to 2018." Journal of Hospitality and Tourism Technology (2019).

[68] Haven, Claire, and David Botterill. "Virtual learning environments in hospitality, leisure, tourism and sport: A review." Journal of Hospitality, Leisure, Sport and Tourism Education 2, no. 1 (2003): 75-92.

[69] Pestek, Almir, and Maida Sarvan. "Virtual reality and modern tourism." Journal of Tourism Futures 7, no. 2 (2020): 245-250.

[70] Loureiro, Sandra Maria Correia. "Virtual reality, augmented reality and tourism experience." The Routledge Handbook of Tourism Experience Management and Marketing (2020): 439-452.

[71] Kononova, Olga, Dmitry Prokudin, and Elena Tupikina. "From e-Tourism to Digital Tourism. Terminologically Review." In SSI, pp. 164-177. 2020.

[72] Batchelor, David, Marc Aurel Schnabel, and Michael Dudding. "Smart heritage: Defining the discourse." Heritage 4, no. 2 (2021): 1005-1015.

[73] Abdelhamid, Tarek Galal. "Digital Techniques for Cultural Heritage and Artifacts Recording." Resourceedings 2, no. 2 (2019): 72-112.

[74] Münster, S. "A survey on topics, researchers and cultures in the field of digital heritage." ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences 4 (2017).

[75] Tammaro, Anna Maria. "Heritage curation in the digital age: Professional challenges and opportunities." International Information & Library Review 48, no. 2 (2016): 122-128.

[76] Loannides, Marinos, and Q. Ewald. "3D research challenges in cultural heritage." Lecture notes in computer science 8355 (2014): 151.

[77] Northern Ireland Toursim Toolkit https://www.tourismni.com/globalassets/business-development/support-by-sector/other-sectors/culture--heritage/toolkit/cultural-heritage-toolkit.pdf.

Using a Fuzzy-Bayesian Approach for Predictive Analysis of Delivery Delay Risk

Ouafae EL Bouhadi[1], Monir Azmani[2], Abdellah Azmani[3], Mouna Atik el ftouh[4]
Intelligent Automation Laboratory, FST of Tangier
Abdelmalek Essaadi University
Tetouan, Morocco

***Abstract*—Although one of the major roles of delivery logistics activities is to ensure a good quality of customer service, certain risks such as damage, delay, return of transported goods occur quite often. This makes risk control and prevention one of the requirements of transport supply chain quality. The article focuses on the analysis of the risk of delay, which is often considered fundamental for the quality of service and as a center of additional costs related to the violation of time windows. Such a risk can harm the image of a supplier, which can even lead to the loss of customers in case of recurrence. The aim of the following case study is the development of a fuzzy-bayesian approach that anticipates, by predictive analysis combining Bayesian networks (BNs) and Fuzzy logic, the possible delays affecting the smooth running of a delivery operation. The results of the implementation of the late delivery risk prediction model are validated by verifying three axioms. In addition, a sensitivity and scenario analysis is performed to validate the model and identify the parameters that have the most adverse impact on the occurrence of such a risk. These results can help carriers/transport providers to minimize potential late deliveries. In addition, the developed model can be used as a basis for different types of predictions in the field of freight transport as well as in other research areas.**

Keywords—Delivery logistics; risk management; predictive analysis; bayesian network; fuzzy logic

I. Introduction

On-time delivery is often considered as a performance indicator in the field of freight transportation [1]. However, any delay in goods delivery to their destination can harm the activities of the actors involved (shipper, carrier, warehouseman, and final customer) and affect their profitability. Therefore, measuring the risk of delay in advance enables reliable decisions to be made when planning deliveries. Consequently, the effect of such a risk can be reduced or even eliminated. This study focuses on the analysis of delivery delay to assist transportation companies in making decisions and optimizing their planning. Despite, the development of a delay risk prediction model is hampered by the unavailability of data. With this in mind, this paper contributes by implementing a fuzzy-bayesian approach to overcome the problem of missing or unavailable data.

The occurrence of delay in freight transport operations is influenced on the one hand by external factors such as road events (accidents, congestion, and weather conditions). On the other hand, by internal factors related to carriers' decisions in terms of resource selection and delivery planning [2]. This paper has studied all of these elements, and it has designed a predictive model integrating the different cause-effect relationships between several factors (internal and external) impacting on-time delivery. For this, it opted for BNs [3] thanks to its advantage of modeling, probabilistic reasoning of uncertain systems [4], and causal analysis.

The article is structured as follows: an overview of the literature related to delivery delay and the application of BNs in the transportation domain is given in Section II. The construction of the fuzzy-Bayesian model and its validation are presented in Section III. Section IV provides a discussion of the results obtained. Finally, the conclusion and further research are covered in Section V.

II. Related Work

In this section, we provide a brief overview of literature regarding our context of study, namely delayed deliveries, and for Bayesian network applications in transportation field.

A. *Delayed Deliveries*

Delayed deliveries mean goods arriving at their destination out of schedule. According to ATRI (American Transportation Research Institute), in 2016, trucking operations in the U.S. experienced about 1.2 billion hours of delays alone due to traffic congestion. This number of delays generates $74.5 billion in additional operational costs [5]. Because there are other causes of delay besides congestion, these additional costs continue to rise.

According to [6] various factors can disrupt delivery reliability, among them personnel issues, vehicle breakdowns, and poor planning. This means that resource selection by checking vehicle conditions and driver performance in addition to the adequacy of planning have a significant effect on speed of service. Furthermore, other external factors such as accessibility to delivery points, accidents, unforeseen events (public works), and methodological conditions [6]–[8] can cause further delays.

The occurrence of delay largely influences [9]–[12]:

- Additional costs: these costs are typically related to a delay penalty, and vehicle operating costs such as maintenance and vehicle rental in the case of private fleet use. As well as driver-related costs, that include salary and benefits. In addition, delay can lengthen the storage period, resulting in inventory holding costs.

- Warehouse productivity: arriving late at the warehouse may result in overlapping deliveries. This will influence the availability of unloading bays and labor as well as overloading the workforce to wait for late deliveries.
- Customer satisfaction: late deliveries can affect the operations of the consignees. This will have a significant impact on the customer relationship. As a result, customers find themselves in a situation of frustration. In case of recurrence, this situation exposes suppliers to the loss of customers.
- Loss of opportunity to consolidate shipments: the occurrence of delays contributes to the lack of opportunity to consolidate shipments due to the uncertainty of meeting delivery deadlines.
- Carrier profitability: this is generally affected by the additional costs incurred, the damage to the carrier's image due to the loss of customers and the extra time generated which reduces the opportunity to make more deliveries.

The concept of predicting the risk of late delivery is generally studied in the context of improving service quality. In the literature, the prediction of such risk is performed using several techniques from artificial intelligence.

Keung et al. [13] have opted for machine learning methods such as KNN, and ANN to predict shipment delays. The authors [14] used Random Forest and SVM for on-time delivery prediction. Berrones-Sanz [15] also proposed a model for on-time delivery prediction using logistic regression. In order to predict on-time delivery violation, [16] relied on ANN machine learning technique. As for [17], they proposed a neural network-based model to anticipate delivery time.

In addition, BNs are among the most popular prediction methods in various research areas [18]. This paper exploits the potential of BNs in risk prediction to anticipate the occurrence of delay in a delivery operation.

B. *Applications of BNs in the Transportation Field*

BNs are applied in several domains: diagnosis (medical and industrial), risk management, spam detection, fraud detection, data mining, and text mining, etc.[19]. In the transportation domain, BNs are used for different types of prevention. Gregoriades and Mouskos [20] used BNs to quantify the risk of accidents in order to locate black areas. The authors [21] developed a bayesian model for the identification of features affecting the safety of motor carriers. Zhu et al. [22] presented a bayesian approach for contextual (e.g., dynamic traffic information) or non-contextual (e.g., instantaneous driving speed) evaluation of driving behavior. As for [23], they modeled drivers' vehicle use behavior according to time of day. The use of BNs also extends to other transportation axes, namely traffic congestion prediction [24], [25] as well as freight demand prediction [26].

III. Modeling the Risk of Late Delivery using a Fuzzy-bayesian Approach

A BN is "a graphical probabilistic model consisting of a set of nodes (variables of interest in the domain) and arcs (causal phenomena). In addition to a set of local probability distributions (network parameters)" [3], [27], [28]. A BN allows modeling the effect of a fact or an uncertain event on another via the representation of the causal relationship between those events [29]. In this case, an arc from X to Y can be interpreted as 'X causes Y'. Fig. 1 shows that the knowledge gained about the overlay of fragile products determines the knowledge about the damage of goods.

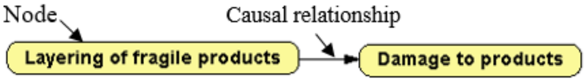


Fig. 1. Example of a Causality Representation.

The study is based on exhaustive bibliographical research, in addition to a survey of experts, in order to identify all the parameters (factors) which can hinder the delivery of a good to its destination in time. As well as the schematization of the dependency relations between these parameters, in order to build the structure of the Bayesian network. The approach followed introduces a methodology for modeling and evaluating the risk of late delivery using BNs by following these steps:

Definition of the BN structure: Determining the relationships between nodes allowed us to design a causal graph based on a three-level architecture.

- The first level represents the input or feeder nodes that have an indirect effect on the risk of delay; they fall into five categories detailed in Table I;
- The second level consists of the intermediate nodes that represent the various intermediate cause factors (direct causes, or cause-effects) that lead to the impact factors encapsulated in the third level (final impacts);
- The third level contains the final impacts that define the factors that directly and negatively influence the arrival of goods at their destination on time.

Generate conditional probabilities of intermediate effects and final impacts: The generation of conditional probability tables will be made using Sugeno's fuzzy inference implementation.

Model validation: This step consists of relying on Bayes theory and posterior probabilities relating to intermediate effects and final impacts, in order to study the effect of a state linked to an input parameter on the envisaged risk. In addition, sensitivity analysis and partial validation are performed to validate the model.

A. *Definition of the BN Structure*

The construction of the Bayesian network architecture can be done in two different ways:

- Objective methods: by using a database to apply the structure's learning methods.

- Subjective methods: by gathering knowledge from experts in the field, through written questionnaires, individual interviews or brainstorming sessions.

The first approach necessitates a large quantity of data. and it may establish dependencies or independencies between some variables that are inconsistent with the experts' opinions [30]. For this reason, researchers prefer the use of the second approach where experts are involved to verify the causal links between the network variables. [31]. In the literature, several researchers have relied on expert knowledge among them [31]–[35].With this in mind, the article used the subjective method.

Based on a survey of experts in the freight transportation field and a literature review of the various factors that drive the occurrence of delay risk, a set of network input parameters are identified and presented in detail in Table I. In addition to the intermediate effects and final impacts presented in Table II. The survey consisted of two questionnaires, the first was conducted to verify and validate the identified variables and the second to establish the causal relationships between the variables.

After identifying the variables (Tables I and II) that will constitute the nodes of the graph and studying the causal relationships between them, the structure of the Bayesian network is developed and illustrated in Fig. 2.

B. *Generation of Conditional Probabilities of Intermediate Effects and Final Impacts*

After building the Bayesian network structure, the next step is to compute the conditional probability tables (CPT) for each variable. These CPTs can be computed based on the knowledge of the experts or using learning algorithms from a database. Although there is no more database in the literature adapted to the parameters identified to build the graph structure, the article is oriented towards the use of subjective methods.

Since the number of conditional probabilities in the developed network is 1032, it is difficult to rely on expert knowledge to evaluate such a large number of probabilities[36]. In fact, in the literature, many researchers have developed models to lower the number of CPTs of a BN. As an example, causal interaction models that have attracted the interest of several researchers such as [37]–[41]. One of the most widely used models is the Noisy-OR model introduced by [39] which allows for the specification of non-deterministic interactions between the parents associated with an effect [42].

In addition, other methods such as fuzzy logic [43] allowing the reduction of numbers of questions asked to experts and the generation of probability tables [2], [44], [45] .With this in mind, the paper relied on this method to first express the experts judgments by fuzzy rules, and then generate the conditional probability tables by a fuzzy inference mechanism. These fuzzy rules are of the type 'If the driver's performance is bad then the occurrence of the accident is high. Here the value of "high" for the accident occurrence is qualitatively represented by a linguistic variable expressed in natural language. For the different rules, the accident occurrence node can be translated into one of the following values: low, medium, high. In addition, for the different nodes of the graph, their influences are revealed by three linguistic values (states) represented in Table III. Since the quantification of these linguistic variables oscillates in interval 0.1, the article opted for the expression of these values in fuzzy form to ascertain the degree of each node's membership in all its fuzzy subsets. As an example, the accident occurrence node is high with 90%, medium with 8% and low with 2%.

The implementation of the approach adopted for generating conditional probability tables is done in three steps:

- Definition of fuzzy variables, their associated linguistic values(variable whose values are qualitative and represent natural language expressions [62] and their membership functions;
- Determination of fuzzy rule bases: a base of "if-then" rules, is used by the "fuzzy inference system" in order to translate the input variables into output [62];
- Development of inference mechanism that forms conclusions based on the fuzzy rules and the input data [63].

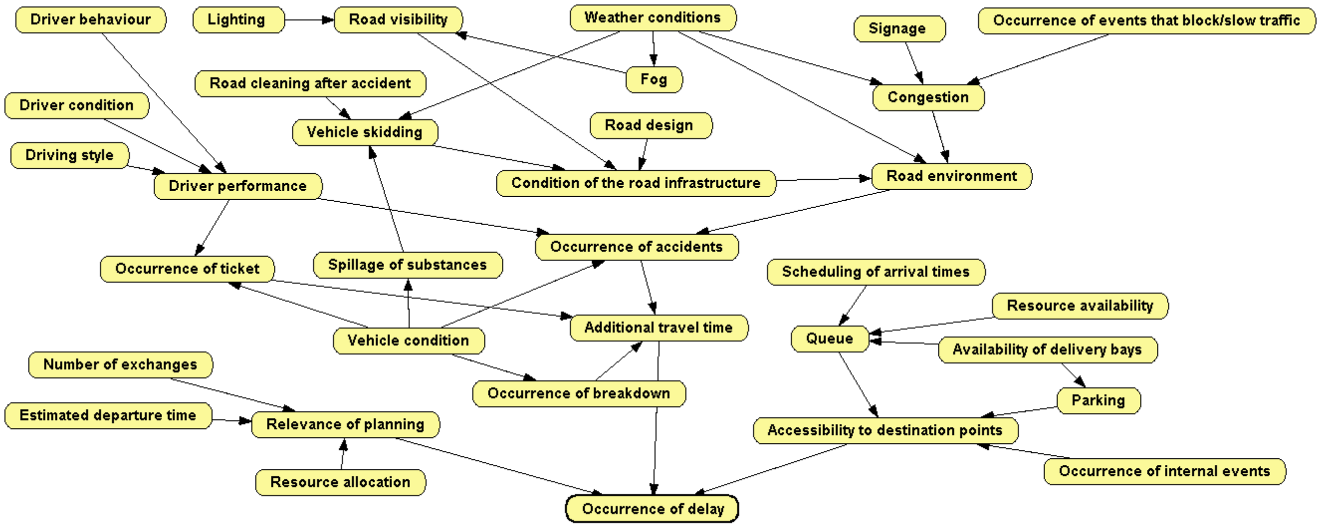


Fig. 2. The Structure of the Bayesian Network Modeling the Risk of Delay of a Delivery.

TABLE I. DESCRIPTION OF THE INPUT PARAMETERS OF THE CAUSAL GRAPH

Variable class	**Variable name**	**Description**
Parameters related to the road	Road design	It refers to the geometric nature of the road, such as the number of lanes and direction of traffic. [46].
	Lighting	Lighting conditions (artificial light, daylight, darkness, twilight) have a considerable effect on the occurrence of road accidents, due to the adaptation of speed with visibility conditions [47].
	Road cleaning after accident	Improper cleaning of accident sites can result in the vehicle skidding due to the presence of body fluids, fuel and other debris.
Traffic parameters	Signage	Traffic signal control considers the control of traffic signals and the existence of stop sign and yield sign [48] .
	Weather conditions	Weather conditions can disrupt driver capabilities, pavement friction, road infrastructure conditions, and vehicle stability and maneuverability, due to reduced visibility, extreme temperatures, precipitation, high winds, and lightning [49]. As a result, they affect traffic demand (carriers postpone or cancel planned delivery operations), traffic safety (accident rates), and traffic flow relationships (changes in the fundamental traffic flow variables, volume, speed, and density influence the capacity of a road system) [50].
	Occurrence of events that block/slow traffic	Traffic can be disrupted by the occurrence of a variety of events, including: Unexpected events such as road accidents, vehicle breakdown in the middle of the road, land subsidence and public works [51]; Irregular social events such as political demonstrations, diplomatic visits [51], sporting events; Regular events such as festive events (religious, national or international holidays) and vacation departure; In addition, the delivery period also influences traffic. Delivery during peak hours can slow down traffic.
Parameters related to personnel (drivers, planners, ...) or planning systems	Driver behavior	Driver behavior is considered one of the major sources of traffic accidents [52]. These behaviors include failure to observe speed limits, safety distance, and poor vehicle handling that manifests itself in hard braking, and hard acceleration, especially when visibility is reduced, and when climbing hills [53].
	Driving style	A driver's habitual driving style that includes calmness behind the wheel, level of attention contribute to traffic accidents [54], [55].
	Driver condition	The poor condition of drivers is often due to fatigue, alcohol consumption, negative emotions (worries and fear of arriving late), drowsiness, headaches, respiratory illnesses, and fever [56].
	Number of exchanges	The increase in the number of exchanges resulting from indirect deliveries leads to more loading/unloading operations [57], and therefore an increase in the risk of delay due to uncontrolled/estimated time at the exchange points.
	Departure time	A travel time estimate is said to be accurate if it could help improve the quality of service by delivering the goods on time [58]. For this, the departure time must be sufficient to cover the planned route within the required time window.
	Resource allocation	Personnel, vehicle condition (breakdown), and delivery points are among the factors that cause delay to occur [6]. Therefore, optimal allocation of resources (vehicles, warehouses, personnel, etc.) to routes helps to reduce the risk of delay.
Vehicle-related parameters	Vehicle condition	Vehicle condition is one of the causes of accidents [59]. Poor physical condition of the vehicle can cause vehicle breakdown and hence the occurrence of delays. In addition, negligence in checking the administrative condition of vehicle can be the cause of vehicle ticketing.
Parameters related to delivery areas (warehouses)	Availability of delivery bays	The lack of adequate delivery bays for different vehicle sizes leads to long queue [60]. In this case, drivers try to park their vehicles far away or illegally near the delivery site. This leads to long queue.
	Resource availability	The lack of human resources in the delivery areas is one of the factors that lead to long queue[60]. Similarly, the lack of material resources (handling equipment, forklifts, etc.) can aggravate this problem.
	Internal events	Internal events such as strikes over working conditions can slow down or block access to delivery areas.
	Scheduling of arrival times	Synchronizing inbound and outbound delivery operations within a warehouse requires reliable arrival times [61]. Lack of scheduling of arrival times (delivery deadlines) with carriers can cause overlapping operations of multiple deliveries, subsequently resulting in long queues.

TABLE II. PRESENTATION OF INTERMEDIATE EFFECTS AND IMPACTS ON DELAY

Intermediate effects		**Final Impacts**
• Road visibility • Vehicle skidding • Condition of the road infrastructure • Spillage of substances • Congestion • Occurrence of ticket	• Road environment condition • Driver performance • Queue • Parking • Occurrence of accidents • Occurrence of breakdown	• Additional travel time • Accessibility to destination points • Relevance of planning • Occurrence of delay

TABLE III. STATES OF THE BAYESIAN NETWORK NODES

Nodes	Linguistic values
Road design	Good, medium, bad
Lighting	Good, medium, bad
Fog	Low, medium, heavy
Road visibility	Good, medium, bad
Condition of the road infrastructure	Good, medium, bad
Road cleaning after accident	Appropriate, medium, inappropriate
Spillage of substances	Low, medium, strong
Vehicle skidding	Low, medium, significant
Weather conditions	Normal, medium, extreme
Occurrence of events that block/slow traffic	Low, medium, high
Signage	Good, medium, bad
Congestion	Low, medium, high
Road environment	Good, medium, bad
Vehicle condition	Good, medium, bad
Driver performance	Good, medium, bad
Driver behavior	Good, medium, bad
Driving style	Good, medium, bad
Driver condition	Good, medium, bad
Occurrence of accidents	Low, medium, high
Occurrence of breakdown	Low, medium, high
Occurrence of ticket	Low, medium, high
Additional travel time	Low, medium, high
Number of exchanges	Low, medium, high
Resource allocation	Good, medium, bad
Relevance of planning	Good, medium, bad
Estimated departure time	Good, medium, bad
Scheduling of arrival times	Good, medium, bad
Occurrence of internal events	Low, medium, high
Resource availability	High, medium, low
Availability of delivery bays	High, medium, low
Queue	Fast, medium, slow
Parking	Very close, close, far
Accessibility to destination points	High, medium, low
Occurrence of delay	Low, medium, high

To better assimilate our proposed approach, we explain CPTs generation for ‘Congestion’ node. In this case, the inference mechanism aims at determining the probabilities of the occurrence of congestion with respect to the states of the nodes: Signage, weather conditions, occurrence of events blocking/disrupting the traffic.

The membership function used for the different nodes of the graph is Gaussian, as it provides less error compared to other triangular and trapezoidal functions [64]. An example of the description of the membership function for the weather variable is provided in Fig. 3.

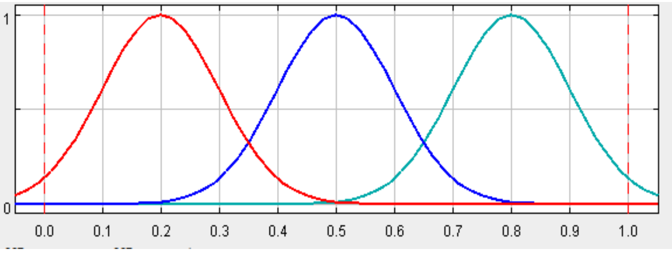


Fig. 3. Membership Functions for the Weather Variable.

After defining the membership functions for the congestion nodes and its antecedents, subsequently, a fuzzy rule base is determined to evaluate the variation of the congestion node with respect to the states of its parent nodes or its causes. This fuzzy rule base is detailed in Table IV.

TABLE IV. FUZZY RULES OF THE 'CONGESTION' NODE WITH ITS PARENT NODES

Rule number	Signage	Weather conditions	Occurrence of events	Congestion
1	Bad	extremes	High	high
2	Bad	extremes	Medium	high
3	Bad	extremes	Low	medium
4	Bad	medium	High	high
5	Bad	medium	Medium	medium
6	Bad	medium	Low	medium
7	Bad	normal	High	medium
8	Bad	normal	Medium	medium
9	Bad	normal	Low	low
10	Medium	extremes	High	high
11	Medium	extremes	Medium	medium
12	Medium	extremes	Low	medium
13	Medium	medium	High	medium
14	Medium	medium	Medium	medium
15	Medium	medium	Low	medium
16	Medium	normal	High	medium
17	Medium	normal	Medium	medium
18	Medium	normal	Low	low
19	Good	extremes	High	high
20	Good	extremes	Medium	medium
21	Good	extremes	Low	medium
22	Good	medium	High	medium
23	Good	medium	Medium	medium
24	Good	medium	Low	low
25	Good	normal	Low	medium
26	Good	normal	Medium	low
27	Good	normal	Low	low

The inference mechanism is triggered by initializing the input variables with precise values representing the peak of the Gaussian distribution and activating a set of fuzzy rules. This mechanism uses the Sugeno inference system. The aim is to identify degrees of membership to each fuzzy subset. The calibration of these degrees provides the CPTs of the BN.

Consider the example of inferring the congestion node knowing that the signage is average, the methodological conditions are normal, and the occurrence of traffic blocking/interfering events is low. The initial values of the input variables represented in rule 18 of Table VI will feed the fuzzy inference system. The latter is implemented with the Fispro software.

The results of the fuzzy inference are illustrated in Fig. 4, thereafter; the different results are aggregated, in order to combine them into a single value for each state. This value is the result of the union of the different conclusions of the rules activated with the max method.

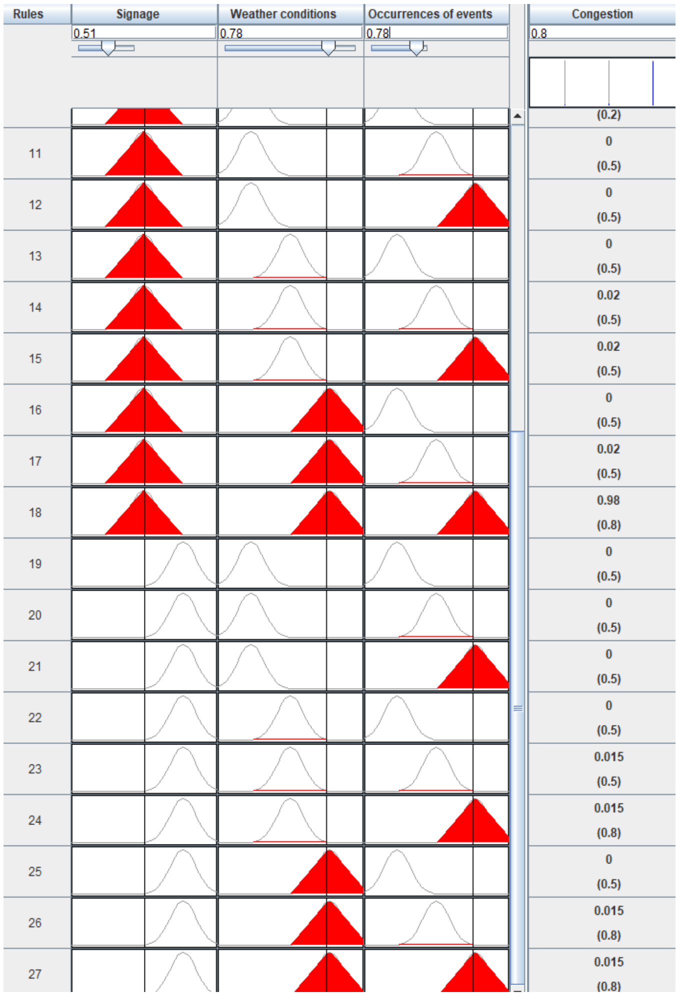


Fig. 4. Fuzzy Inference of the Variable 'Congestion.

The conclusions of the activated rules of rule 18 are summarized in Table V. Thus, the low value of congestion is 0.98, and the average value is 0.020. Now, since each possibility of a fuzzy subset must be greater than zero, a value of 0.001 is given to the null probability, in this case the high value is 0.001.

Congestion (low) = max (0.98, 0.015, 0.008) = 0.98.

Congestion (medium) = max (0.015, 0.008, 0.020) = 0.020.

Congestion (high) = 0.001.

The conditional probabilities for the different states of the variable 'congestion' of the rule 18 is calculated in the following way:

P (Congestion = low | Occurrence events = low, Weather = normal and Signage = medium) = 0.98/(0.98+0.020+0.001)=0.979.

P (Congestion = medium | Occurrence of events = low, Weather = normal and Signage = medium) = 0.020/(0.98+0.020+0.001)=0.019.

P (Congestion = high | Occurrence of events = low, Weather = normal and Signage = medium) = 0.001/(0.98+0.020+0.001)=0.001.

TABLE V. DEGREE OF MEMBERSHIP FOR EACH FUZZY SUBSET OF THE VARIABLE 'CONGESTION

Rule activated	**Language value of the output variable**	**Degree of membership**
R27	Low	0.015
R26	Low	0.015
R24	Low	0.015
R23	Medium	0.015
R18	Low	0.98
R17	Medium	0.020
R15	Medium	0.020
R14	medium	0.020
R9	Low	0.008
R8	medium	0.008
R6	medium	0.008

C. *Anticipation of Scenarios and Interpretation of Results*

After constructing the Bayesian network using Open Markov tool, it is used to deduce the probabilities of certain events by setting evidences (states) for certain nodes and study their effects through the propagation of their probabilities on the child nodes. At this level, the input parameters' impact on the occurrence of delay is studied through four scenarios listed as follows:

- Scenario 1: favorable road environment and delivery areas input parameters, unfavorable transport company input parameters.
- Scenario 2: unfavorable road environment and delivery areas input parameters, favorable transport company input parameters.

- Scenario 3: favorable road environment and delivery areas input parameters, favorable transport company input parameters.
- Scenario 4: unfavorable road environment and delivery areas input parameters, unfavorable transport company input parameters.

The input parameters related to the transport company correspond to those related to the vehicles, the staff (drivers, planners ...) and the planning systems.

The input parameters related to the road environment and delivery areas correspond to those related to the road, traffic and warehouses.

The four scenarios corresponding to different configurations of the input node states are detailed in Table VII.

After feeding the BN with the states of each scenario, the inference mechanism provides probability propagation over intermediate effects and final impacts to quantify delay occurrence. Table VIII, shows the probability distribution for some nodes in the network.

TABLE VI. CONDITIONAL PROBABILITY TABLE OF THE 'CONGESTION' NODE

					Conditional probability		
Rule number	**Signage**	**Weather conditions**	**Occurrences of events**	**Congestion**	**Low**	**medium**	**high**
1	bad 0.22	extremes 0.18	high 0.21	high	0.001	0.015	0.984
2	bad 0.22	extremes 0.18	medium 0.52	high	0.001	0.020	0.979
3	bad 0.22	extremes 0.18	low 0.78	medium	0.001	0.979	0.020
4	bad 0.22	medium 0.5	high 0.21	high	0.001	0.020	0.979
5	bad 0.22	medium 0.5	medium 0.52	medium	0.001	0.988	0.011
…	…	…	…	…	…	…	…
18	medium 0.51	normal 0.78	low 0.78	low	0.979	0.020	0.001
19	good 0.81	extremes 0.18	high 0.21	high	0.001	0.015	0.984
20	good 0.81	extremes 0.18	medium 0.52	medium	0.006	0.988	0.006
…	…	…	…	…	…	…	…
25	good 0.81	normal 0.78	high 0.21	medium	0.015	0.984	0.001
26	good 0.81	normal 0.78	medium 0.52	low	0.979	0.020	0.001
27	good 0.81	normal 0.78	low 0.78	low	0.979	0.020	0.001

TABLE VII. DESCRIPTION OF THE SCENARIOS ACCORDING TO THE VALUES OF THE INPUT PARAMETERS

Input parameter	**Value**			
	Scenario 1	***Scenario 2***	***Scenario 3***	***Scenario 4***
Road environment and delivery areas parameters				
Road design	good	bad	good	bad
Lighting	good	bad	good	bad
Post-accident road cleaning	appropriate	inappropriate	appropriate	inappropriate
Signage	good	bad	good	bad
Weather conditions	normal	extremes	normal	extreme
Occurrence of events that block/impede traffic	low	high	low	high
Availability of delivery bays	high	low	high	low
Resources Availability	high	low	high	low
Occurrence of internal events	low	high	low	high
Scheduling of arrival times	good	bad	good	bad
Transport company parameters				
Driving style	bad	good	good	bad
Driver condition	bad	good	good	bad
Driver behavior	bad	good	good	bad
Vehicle condition	bad	good	good	bad
Resource allocation	bad	good	good	bad
Estimated departure time	bad	good	good	bad
Number of exchanges	high	Low	low	high

TABLE VIII. PROBABILITY DISTRIBUTION FOR BAYESIAN NETWORK NODES

Node	State	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Congestion	Low	0.9790	0.0010	0.9790	0.0010
	Medium	0.0020	0.0150	0.0020	0.0150
	High	0.0010	0.9840	0.0010	0.9840
Occurrence of a breakdown	high	0.9790	0.0010	0.0010	0.9790
	Medium	0.0200	0.0020	0.0020	0.0174
	Low	0.0027	0.9790	0.9790	0.0010
Occurrence of an accident	Low	0.0022	0.0012	0.9574	0.0010
	Medium	0.0418	0.9574	0.0416	0.0173
	high	0.9559	0.0214	0.0010	0.9817
	Low	0.0209	0.9166	0.9166	0.0052
Occurrence of ticket	Medium	0.0052	0.0592	0,0592	0.0209
	high	0.9738	0.0242	0.0242	0.9738
Accessibility to destination points	High	0.9635	0.0010	0.9635	0.0010
	Medium	0.0355	0.0155	0.0355	0.0155
	Low	0.0010	0.9835	0.0010	0.9835
Delay	Low	0.0017	0.0050	0.9718	0.0010
	Medium	0.0861	0.9624	0.0272	0.0160
	High	0.9121	0.0327	0.0010	0.9830

In the case of the first scenario, the occurrence of congestion is low with a probability of 97.9%. Concerning the occurrence of a breakdown and ticket, and accident are important with respectively 97.9% and 97.3% and 95.5%. As for the accessibility to the destination, point tends to be high with 96.3%. Therefore, the probability of the occurrence of delay is important with a value of 91.1%.

For the second scenario, the probability of congestion occurrence is high with a probability of 98.4%. Regarding the occurrence of accident is more likely to be 'medium' with a probability of 95.7%, as for the occurrence of breakdown and ticket are low with respectively 97.9% and 91.6%, In addition, the accessibility to the destination point tends to be low with 98.3% which leads to an average risk of occurrence of delay of 96.2%

For the third scenario, the inference model predicts a low probability of congestion occurrence with a rate of 97.9%. Also, the risks related to the occurrence of an accident, breakdown and a ticket are low with respectively 95.7%, 97.9% and 91.6%. In addition, the accessibility to the destination point tends to be high with 96.3%. Therefore, the probability of delay occurrence is very low with 97.1%.

For the fourth scenario, congestion tends to be high with a rate of 98.4%. Also, the risks of the occurrence of an accident, a breakdown or a ticket are high with respectively 98.1%, 97.9% and 97.3%. As for the accessibility to the destination, point is low with a percentage of 98.3%. Hence the highest probability of the occurrence of delay with 98.3%.

Based on the results of the inference of the first two scenarios, we can conclude that the parameters related to the transport company have more impact on the occurrence of delivery delay than those related to road and delivery environment.

D. Sensitivity Analysis

The sensitivity analysis identifies the factors ranked according to those that have more impact on the probability of a node [65]. The Fig. 5 shows the tornado diagram [66] of the "delay occurrence" sensitivity analysis. According to the figure, variables such as vehicle condition, resources availability, estimated departure time, number of exchanges, and resource allocation contribute more to the occurrence of late deliveries. Poor vehicle condition is the main parameter influencing the risk of delay. This can be justified by its resulting events such as the occurrence of breakdowns and accidents, as these have a significant impact on the occurrence of such risk.

E. Partial Validation

In order to validate the proposed model, the paper uses partial validation allowing the verification of three axioms presented by [67] and opted by various researchers such as [68], [69]: (1) The occurrence of change (increase/decrease) in the probability of the parent node thus changes the probability of the child node; (2) The values of the child nodes must be consistently affected by the changes made to the probability distributions relative to the parent node; In the case of a node with more than one parent (e.g., x and y), the overall effect of x and y must be greater than the individual effects of parent x or parent y.

For example, when the probability of "road design = good" increases from 70% to 75%, the probability of "road infrastructure condition = good" and "road environment = good" increases from 77% to 80% and 84% to 86% respectively. In light of this variation, when the probability of "Occurrence of events that block/slow traffic=low" increases from 74% to 78%, the probability of "congestion=low" and "road environment=good" increases from 78% to 82% and 87% to 90% respectively, which is consistent with the axiom. Similarly for the other nodes of the graph.

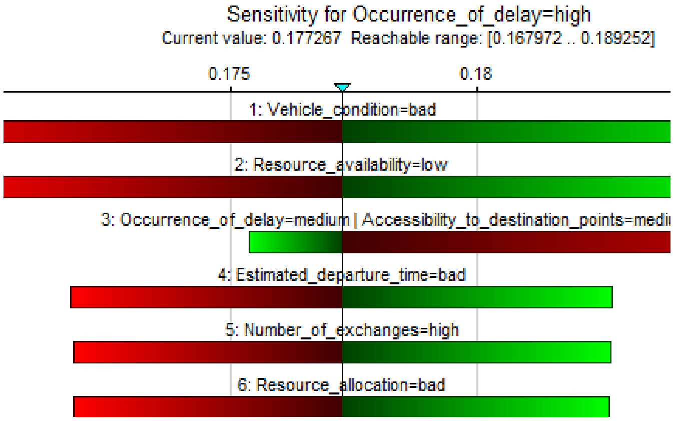


Fig. 5. Sensitivity Analysis for Occurnce of Delay.

IV. Discussion

In this paper, an analysis of potential delivery delays is developed using a Bayesian fuzzy model. BNs have been used in various existing studies to analyze the risk of delay in different areas such as maritime transport[45], [70], rail transport [71] and air transport [72]. Compared to the existing works, the developed model focuses on the occurrence of delay in the road freight transport domain. This model consists of many internal and external factors that cause delay to occur. Sensitivity analysis and interpretation of the results of four scenarios show that the internal factors related to the transport companies have a stronger effect on the occurrence of late deliveries than the external factors. This means that the transport company decisions in terms of resource selection, in addition to planning relevance, have a considerable effect on delivery reliability than road and delivery events such as congestion, weather conditions and availability of delivery bays. Therefore, optimized resource (physical and material) allocation to the right routes and a smart routing planning design plays an important role in ensuring on-time deliveries. In order to validate the results of this model, an example test of three axioms is performed.

V. Conclusion

The respect of delivery deadlines is crucial to ensure the quality of logistics services. Unfortunately, it often happens that this deadline is not respected, leading to a series of more or less unfortunate consequences. This article focuses on predicting the risks of delays in deliveries with the aim of anticipating them in order to either avoid them or be better prepared to deal with them and reduce the impact of poor quality of service. To do this, the article proposes a fuzzy Bayesian model combining the Bayesian approach and fuzzy logic in order to monitor the occurrence of delivery delays. This model is based on a set of factors causing the delay of deliveries and their causal relationships represented by a causal graph. Fuzzy logic intervenes by the generation of fuzzy rules based on conditional probability tables. Such a model is particularly effective, especially for the type of problem dealt with through this article. It is positioned as an excellent alternative to deep learning for making predictions in the absence of massive data.

One of the limitations of the proposed model is the identification of all the factors that cause delivery delays, the lack of integration of all these factors can affect the effectiveness of this model in correctly predicting possible delays.

The generalization of the model to all risks that can degrade the quality of delivery services such as damage, theft and loss, is a potential perspective to the work presented here. Similarly, a comparative study between a Bayesian-fuzzy model and a model derived from deep-learning around the prediction of delivery risks and the management of the crises that follow are very promising avenues of research.

Acknowledgment

This research is supported by the Ministry of Higher Education, Scientific Research and Innovation, the Digital Development Agency (DDA) and the National Center for Scientific and Technical Research (CNRST) of Morocco (Smart DLSP Project - AL KHAWARIZMI IA-PROGRAM).

References

[1] R. D. Alcoba et K. W. Ohlund, « Predicting On-time Delivery in the Trucking Industry », p. 4, 2017.

[2] M. ATIK EL FTOUH, « Modélisation d'un smart écosystème numérique pour une gestion optimisée et collaborative du transport urbain des marchandises, agrégeant pour l'aide à la décision plusieurs méthodes et techniques de l'intelligence artificielle », 2020.

[3] J. Pearl, Bayesian networks: A model of self-activated memory for evidential reasoning. UCLA, Computer Science Dept, 1985.

[4] J. Chen et al., « shinyBN: an online application for interactive Bayesian network inference and visualization », BMC Bioinformatics, vol. 20, no 1, p. 711, déc. 2019, doi: 10.1186/s12859-019-3309-0.

[5] A. Hooper, « Cost of Congestion to the Trucking Industry: 2018 Update », p. 41, 2018.

[6] A. Mckinnon, J. Edwards, M. Piecyk, et A. Palmer, « Traffic congestion, reliability and logistical performance: A multi-sectoral assessment », International Journal of Logistics-research and Applications - INT J LOGIST-RES APPL, vol. 12, oct. 2009, doi: 10.1080/1367556090 3181519.

[7] O. Korbaa, A. Boufaied, et C. Lajimi, « Assessing and modelling transport delays risk in supply chains », IJAOM, vol. 9, no 4, p. 225, 2017, doi: 10.1504/IJAOM.2017.10010983.

[8] T. Bektas, « Freight Transport and Distribution », p. 286, 2017.

[9] H. Chevroton, Y. Kergosien, et J.-C. Billaut, « Problème de tournée avec pénalités de départ et de retard », p. 9, 2019.

[10] A. Fowkes, P. E. Firmin, G. Tweddle, et T. Whiteing, « How highly does the freight transport industry value journey time reliability - and for what reasons? », International Journal of Logistics, vol. 7, mars 2004, doi: 10.1080/13675560310001619259.

[11] Q. Gong, Q. Miao, B. X. Wang, et T. M. Adams, « Assessing Public Benefits and Costs of Freight Transportation Projects: Measuring Shippers' Value of Delay on the Freight System », Art. no CFIRE 04-14, juill. 2012.

[12] V. Lukinskiy, N. Pletneva, V. Gorshkov, et P. Druzhinin, « Application of the Logistics "Just in Time" Concept to Improve the Road Safety », Transportation Research Procedia, vol. 20, p. 418-424, 2017, doi: 10.1016/j.trpro.2017.01.068.

[13] K. L. Keung, C. Lee, et Y. Yiu, A Machine Learning Predictive Model for Shipment Delay and Demand Forecasting for Warehouses and Sales Data. 2021, p. 1014. doi: 10.1109/IEEM50564.2021.9672946.

[14] A. Durjoy, A. Bristy, et M. Hasan, E-commerce shipping: Prediction of on time delivery of products using Data Mining. 2021.

[15] L. Berrones-Sanz, « PREDICTING ON-TIME DELIVERIES IN TRUCKING: A MODEL BASED ON THE WORKING CONDITIONS OF DRIVERS », AT, vol. 7, no 2, p. 47-53, juin 2021, doi: 10.22306/atec.v7i2.105.

[16] R. Shabbar et A. Kasasbeh, Proactive Event Management using ANN with PSO Prediction in Transport Processes. 2017.

[17] A. M. Ghaithan, I. Alarfaj, A. Mohammed, et O. Qasim, « A neural network-based model for estimating the delivery time of oxygen gas cylinders during COVID-19 pandemic », Neural Comput & Applic, vol. 34, no 13, p. 11213-11231, juill. 2022.

[18] S. Borujeni, N. Nguyen, S. Nannapaneni, E. Behrman, et J. Steck, Experimental evaluation of quantum Bayesian networks on IBM QX hardware. 2020.

[19] O. PARENT et J. EUSTACHE, « Les Réseaux Bayésiens A la recherche de la vérité ». Université Claude Bernard Lyon, 2007.

[20] A. Gregoriades et K. C. Mouskos, « Black spots identification through a Bayesian Networks quantification of accident risk index », Transportation Research Part C: Emerging Technologies, vol. 28, p. 28-43, mars 2013, doi: 10.1016/j.trc.2012.12.008.

[21] S. Hwang, L. N. Boyle, et A. G. Banerjee, « Identifying characteristics that impact motor carrier safety using Bayesian networks », Accident Analysis & Prevention, vol. 128, p. 40-45, juill. 2019, doi: 10.1016/j.aap.2019.03.004.

[22] X. Zhu, Y. Yuan, X. Hu, Y.-C. Chiu, et Y.-L. Ma, « A Bayesian Network model for contextual versus non-contextual driving behavior assessment », Transportation Research Part C Emerging Technologies, vol. 81, p. 172-187, août 2017, doi: 10.1016/j.trc.2017.05.015.

[23] D. Li, T. Miwa, et T. Morikawa, « Modeling time-of-day car use behavior: A Bayesian network approach », Transportation Research Part D: Transport and Environment, vol. 47, p. 54-66, août 2016, doi: 10.1016/j.trd.2016.04.011.

[24] T. Afrin et N. Yodo, « A probabilistic estimation of traffic congestion using Bayesian network », Measurement, vol. 174, p. 109051, avr. 2021, doi: 10.1016/j.measurement.2021.109051.

[25] Y. CAO, C. WANG, Y. YANG, J. XU, et Y. GAO, « Multicause Automatic Real-time Identification of Urban Road Traffic Congestion Based on Bayesian Network », 30 décembre 2020.

[26] M. Petri, G. Fusco, et A. Pratelli, « A New Data-Driven Approach to Forecast Freight Transport Demand », juin 2014, vol. 8582, p. 401-416. doi: 10.1007/978-3-319-09147-1_29.

[27] C. J. Butz, S. Hua, J. Chen, et H. Yao, « A simple graphical approach for understanding probabilistic inference in Bayesian networks », Information Sciences, vol. 179, no 6, p. 699-716, mars 2009, doi: 10.1016/j.ins.2008.10.036.

[28] S. Verron, « Diagnostic et surveillance des processus complexes par réseaux bayésiens », 2007.

[29] T. D. Nielsen et F. V. JENSEN, Bayesian Networks and Decision Graphs. Springer Science & Business Media, 2009.

[30] F. Ghasemi, O. Kalatpour, A. Moghimbeigi, et I. Mohammadfam, « Selecting Strategies to Reduce High-Risk Unsafe Work Behaviors Using the Safety Behavior Sampling Technique and Bayesian Network Analysis », J Res Health Sci, vol. 17, no 1, p. 372, mars 2017.

[31] I. Mohammadfam, F. Ghasemi, O. Kalatpour, et A. Moghimbeigi, « Constructing a Bayesian network model for improving safety behavior of employees at workplaces », Applied Ergonomics, vol. 58, p. 35-47, janv. 2017, doi: 10.1016/j.apergo.2016.05.006.

[32] S. A. Abdulkareem, Y. T. Mustafa, E.-W. Augustijn, et T. Filatova, « Bayesian networks for spatial learning: a workflow on using limited survey data for intelligent learning in spatial agent-based models », Geoinformatica, vol. 23, no 2, p. 243-268, avr. 2019, doi: 10.1007/s10707-019-00347-0.

[33] R. Kaya et B. Yet, « Building Bayesian networks based on DEMATEL for multiple criteria decision problems: A supplier selection case study », Expert Systems with Applications, vol. 134, p. 234-248, nov. 2019, doi: 10.1016/j.eswa.2019.05.053.

[34] S. Zhu, X. Cai, J. Lu, et Y. Peng, « Analysis of factors affecting serious multi-fatality crashes in China based on Bayesian network structure », Advances in Mechanical Engineering, vol. 9, p. 168781401770414, juin 2017, doi: 10.1177/1687814017704145.

[35] P. Huang et al., « A Bayesian network model to predict the effects of interruptions on train operations », Transportation Research Part C: Emerging Technologies, vol. 114, p. 338-358, mai 2020, doi: 10.1016/j.trc.2020.02.021.

[36] S. Renooij, « Probability elicitation for belief networks: issues to consider », The Knowledge Engineering Review, 2001, doi: 10.1017/S0269888901000145.

[37] D. Heckerman et J. Breese, « Causal independence for probability assessment and inference using Bayesian networks », Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, vol. 26, p. 826-831, déc. 1996, doi: 10.1109/3468.541341.

[38] P. J. Lucas, « Bayesian network modelling through qualitative patterns », Artificial Intelligence, vol. 163, p. 233-263, déc. 2003, doi: 10.1016/j.artint.2004.10.011.

[39] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988.

[40] S. P. D. Woudenberg, L. C. van der Gaag, et C. M. A. Rademaker, « An intercausal cancellation model for Bayesian-network engineering », International Journal of Approximate Reasoning, vol. 63, p. 32-47, août 2015, doi: 10.1016/j.ijar.2015.05.011.

[41] Y. Xiang et N. Jia, « Modeling Causal Reinforcement and Undermining for Efficient CPT Elicitation », IEEE Transactions on Knowledge and Data Engineering, vol. 19, no 12, p. 1708-1718, déc. 2007, doi: 10.1109/TKDE.2007.190659.

[42] S. Srinivas, « A Generalization of the Noisy-Or Model », in Uncertainty in Artificial Intelligence, Elsevier, 1993, p. 208-215. doi: 10.1016/B978-1-4832-1451-1.50030-5.

[43] L. A. Zadeh, « Fuzzy sets », Information and Control, vol. 8, no 3, p. 338-353, juin 1965, doi: 10.1016/S0019-9958(65)90241-X.

[44] C. LAMAAKCHAOUI, « Modélisation d’un Ecosystème Numérique Favorisant la Transformation Digitale et l’Aide à la Décision Marketing et agrégeant plusieurs Concepts et Méthodes de l’Intelligence Artificielle, un Système Multi-Agent et son Ontologie », ABDELMALEK ESSAADI, 2018.

[45] N. H. M. Salleh, R. Riahi, Z. Yang, et J. Wang, « Predicting a Containership’s Arrival Punctuality in Liner Operations by Using a Fuzzy Rule-Based Bayesian Network (FRBBN) », The Asian Journal of Shipping and Logistics, vol. 33, p. 95-104, juill. 2017.

[46] P. Álvarez, M. A. Fernández, A. Gordaliza, A. Mansilla, et A. Molinero, « Geometric road design factors affecting the risk of urban run-off crashes. A case-control study », PLOS ONE, vol. 15, no 6, p. e0234564, juin 2020, doi: 10.1371/journal.pone.0234564.

[47] A. K. Jägerbrand et J. Sjöbergh, « Effects of weather conditions, light conditions, and road lighting on vehicle speed », SpringerPlus, vol. 5, no 1, p. 505, avr. 2016, doi: 10.1186/s40064-016-2124-6.

[48] X. Zou et W. Yue, « A Bayesian Network Approach to Causation Analysis of Road Accidents Using Netica », Journal of Advanced Transportation, vol. 2017, déc. 2017, doi: 10.1155/2017/2525481.

[49] P. Pisano et L. C. Goodwin, « Surface Transportation Weather Applications », p. 11, 2002.

[50] T. H. Maze, M. Agarwal, et G. Burchett, « Whether Weather Matters to Traffic Demand, Traffic Safety, and Traffic Operations and Flow », Transportation Research Record, vol. 1948, no 1, p. 170-176, janv. 2006, doi: 10.1177/0361198106194800119.

[51] M. Rao et K. R. Rao, « Measuring Urban Traffic Congestion – A Review », International Journal for Traffic and Transport Engineering, vol. 2, p. 286-305, déc. 2012, doi: 10.7708/ijtte.2012.2(4).01.

[52] E. Taniguchi, R. G. Thompson, et T. Yamada, « Concepts and Visions for Urban Transport and Logistics Relating to Human Security », in Urban Transportation and Logistics, CRC Press, 2014.

[53] T. F. Fwa, « Transport and Logistics in Asian Cities », in Urban Transportation and Logistics, CRC Press, 2014.

[54] K. Bucsuházy, E. Matuchová, R. Zůvala, P. Moravcová, M. Kostíková, et R. Mikulec, « Human factors contributing to the road traffic accident occurrence », Transportation Research Procedia, vol. 45, p. 555-561, janv. 2020, doi: 10.1016/j.trpro.2020.03.057.

[55] L. Eboli, G. Mazzulla, et G. Pungillo, « How to define the accident risk level of car drivers by combining objective and subjective measures of driving style », Transportation Research Part F: Traffic Psychology and Behaviour, vol. 49, p. 29-38, août 2017, doi: 10.1016/j.trf.2017.06.004.

[56] F. Alonso, C. Esteban, J. Sanmartín, et S. A. Useche, « Reported prevalence of health conditions that affect drivers », Cogent Medicine, vol. 4, no 1, p. 1303920, janv. 2017, doi: 10.1080/2331205X.2017 .1303920.

[57] E. Taniguchi et R. G. Thompson, Éd., City Logistics: Mapping The Future. Boca Raton: CRC Press, 2014. doi: 10.1201/b17715.

[58] H.-E. LIN et R. Zito, « A review of travel-time prediction in transport and logistics », Proceedings of the Eastern Asia Society for Transportation Studies, vol. 5, janv. 2005.

[59] V. Ratanavaraha et S. Suangka, « Impacts of accident severity factors and loss values of crashes on expressways in Thailand », IATSS Research, vol. 37, no 2, p. 130-136, mars 2014, doi: 10.1016/j.iatssr.2013.07.001.

[60] J. G. V. Vieira, J. C. Fransoo, et C. D. Carvalho, « Freight distribution in megacities: Perspectives of shippers, logistics service providers and carriers », Journal of Transport Geography, vol. 46, p. 46-54, juin 2015, doi: 10.1016/j.jtrangeo.2015.05.007.

[61] S. Van der Spoel, C. Amrit, et J. Hillegersberg, « Predictive Analytics for Truck Arrival Time Estimation: A Field Study at a European

Distribution Center », International Journal of Production Research, vol. In Press, p. 1-21, janv. 2016, doi: 10.1080/00207543.2015.1064183.

[62] H. Sattar et al., « Smart Wound Hydration Monitoring Using Biosensors and Fuzzy Inference System », Wireless Communications and Mobile Computing, vol. 2019, p. e8059629, déc. 2019, doi: 10.1155/2019/8059629.

[63] M.-D. Pop, O. Proștean, T.-M. David, et G. Proștean, « Hybrid Solution Combining Kalman Filtering with Takagi–Sugeno Fuzzy Inference System for Online Car-Following Model Calibration », Sensors, vol. 20, no 19, Art. no 19, janv. 2020, doi: 10.3390/s20195539.

[64] S. Mandal, J. Choudhury, et S. Bhadra Chaudhuri, « In Search of Suitable Fuzzy Membership Function in Prediction of Time Series Data », International Journal of Computer Science Issues, vol. 9, mai 2012.

[65] M. Aydin, E. Akyuz, O. Turan, et O. Arslan, « Validation of risk analysis for ship collision in narrow waters by using fuzzy Bayesian networks approach », Ocean Engineering, vol. 231, p. 108973, juill. 2021, doi: 10.1016/j.oceaneng.2021.108973.

[66] D. S. N. A. Shamsuddin et al., « Dynamic Hazard Identification on SOFC system using Bayesian Network », International Journal of Integrated Engineering, vol. 14, no 2, Art. no 2, juin 2022.

[67] B. Jones, I. Jenkinson, Z. Yang, et J. Wang, « The use of Bayesian network modelling for maintenance planning in a manufacturing industry », Reliability Engineering & System Safety, vol. 95, no 3, p. 267-277, mars 2010, doi: 10.1016/j.ress.2009.10.007.

[68] B. Kamal et M. Aydın, « Application of fuzzy Bayesian approach on bankruptcy causes for container liner industry », Research in Transportation Business & Management, vol. 43, p. 100769, juin 2022.

[69] Z. Liu, Q. Ma, B. Cai, X. Shi, C. Zheng, et Y. Liu, « Risk coupling analysis of subsea blowout accidents based on dynamic Bayesian network and NK model », Reliability Engineering & System Safety, vol. 218, p. 108160, févr. 2022, doi: 10.1016/j.ress.2021.108160.

[70] F. Goerlandt et S. Islam, « A Bayesian Network risk model for estimating coastal maritime transportation delays following an earthquake in British Columbia », Reliability Engineering & System Safety, vol. 214, p. 107708, oct. 2021, doi: 10.1016/j.ress.2021.107708.

[71] F. Corman et P. Kecman, « Stochastic prediction of train delays in real-time using Bayesian networks », Transportation Research Part C: Emerging Technologies, vol. 95, p. 599-615, oct. 2018, doi: 10.1016/j.trc.2018.08.003.

[72] C.-L. Wu et K. Law, « Modelling the delay propagation effects of multiple resource connections in an airline network using a Bayesian network model », Transportation Research Part E: Logistics and Transportation Review, vol. 122, p. 62-77, févr. 2019.

Mining Educational Data to Analyze the Student's Performance in TOEFL iBT Reading, Listening and Writing Scores

Khaled M. Hassan[1]
Demonstrator at ISA Information System, International Smart Association, International Smart Association, ISA, Nasr City, Egypt

Mohammed Helmy Khafagy[2]
Professor of Computer Science
Computer Science Department
Faculty of Computers and Information, Fayoum University
Fayoum

Mostafa Thabet[3]
Lecturer of Information System,
Information System Department
Faculty of Computers and Information, Fayoum University
Fayoum

***Abstract*—Student scores in TOEFL IBT reading, listening, and writing may reveal weaknesses and deficiencies in educational institutions. Traditional approaches and evaluations are unable to disclose the significant information hidden inside the student's TOEFL score. As a result, data mining approaches are widely used in a wide range of fields, particularly education, where it is recognized as Educational Data Mining (EDM). Educational data mining is a prototype for handling research issues in student data which can be used to investigate previously undetected relationships in a huge database of students. This study used the EDM to define the numerous factors that influence students' achievement and to create observations using advanced algorithms. The present study explored the relationship among university students' previous academic experience, gender, student place and their current course attendance within a sample of 473 (225 male and 248 female). Educational specialists must find out the causes of student dropout in TOEFL scores. The results of the study showed that the model could be suitable for investigation of important aspects of student outcomes, the present research was supposed to use the statistical package for social sciences (SPSS V26) for both descriptive and inferential statistics and multiple linear regressions to improve their scores.**

Keywords*—Educational data mining; students score; linear regression; TOEFL; Statistics

I. Introduction

Over the last decade, test developers and experts have fixated much of their time and focus on developing a theoretical view of language ability in order to understand better the nature of language proficiency, as well as developing and applying more sophisticated statistical tools to analyze language tests and test takers' performance in order to best tap these issues[1]. However, language testing research shows that language aptitude is not the only factor influencing test takers' performance. Almost all screening processes in academic environments, from seeking college admission to applying for an exchange student programmer, require the applicant to present TOEFL iBT or other Standard English language test scores.

The TOEFL iBT (Test of English as a Foreign Language) Language testing is largely concerned with whether the results clearly effectively reflect test takers' underlying ability in a certain area in a given testing setting [2]. After graduation, English proficiency is necessary for developing career options and attaining aspirational goals in the workplace [3]. The Educational Testing Service (ETS) commissioned a recent survey study and found a high link between high English proficiency and the income of young professionals (full-time workers in their 20s or 30s) across all major industries. This higher income allows them to put more money into improving their English abilities, which are "a vital instrument for success in today's world". Test-takers personality factors to the testing scenario, such as education level, Gender, and place, can all affect their performance [4]. But these construct-irrelevant elements are regarded as potential causes of test bias, which might cause the acquired results to be unrepresentative of the underlying skill that a language test is attempting to assess. As a result, a thorough assessment of the likely effects of such factors is worthwhile.

Taking these factors into account and the popularity of the TOEFL iBT as a proficiency exam worldwide, this study aims to determine the future effects of test education level, Gender, and place on TOEFL iBT listening reading and writing results.

II. Literature Survey

Test fairness is a challenging topic in the literature when it comes to language testing. Debates about test fairness aim to create tests free of discrimination and contribute to testing equity [5, 6]. When students with the same language ability perform differently on a test, it may be called discriminatory. When the substance of the test is discriminatory to test takers from certain groups, other criteria such as education level, Gender, and test place play a factor. The test's requirements may have different impacts on test takers from different groups; test taker factors such as education level, Gender and place can all contribute to test bias.

These factors can impact a test's validity and lead to measurement mistakes. As a consequence, in the design and development of language exams decreasing the impact of these factors that are not part of the language competence is a top objective [7].

The association between TOEFL score and GPA was shown to be positive and statistically significant; however, it was less for engineering students than for students in other professions and for engineering courses than for non-engineering courses. In logistic regressions of CAE pass rate and graduation rate, the TOEFL score was also statistically significant, showing an increased probability of success with a higher TOEFL score. However, model goodness-of-fit values were low, showing that many students defied overall trends in their performance [8].

Accord to the previous survey, a mixed ANOVA was used to answer the following study question: Is there a significant difference between pre and post TOEFL test scores for male and female students? Is there an interaction between male and female students' pre and post TOEFL test scores? According to those findings, there was a substantial change between pre and post TOEFL exam scores, but no significant variation between genders. Furthermore, no correlation was found between male and female students' pre and post TOEFL test scores [9].

In agreement with the past research, there was a relationship between overseas students' academic performance and their language skills, academic self-concept and other factors that influence academic achievement. The research looked at first-year international students enrolled in undergraduate business programs at a Canadian English-medium institution. The following data was gathered on the students: grades in degree program courses, annual GPA, and EPT scores (including sub scores).

Students also filled out an academic self-concept measure. In addition, instructors in two obligatory first-year business courses were interviewed regarding the academic and linguistic requirements in their courses and the profile of successful students to acquire additional information about success in first-year business courses [10].

In the other side the purpose of this study was to determine whether there was a significant difference in the capacity of male and female students to respond to factual and vocabulary-in-context questions on the TOEFL-like reading comprehension test. The results of reading comprehension tests taken from twenty-one male and twenty-one female students in the English Education Program were used for secondary data analysis. Through the use of random sampling, samples were chosen. Utilizing an independent sample t-test, data were evaluated [11].

On the other hand in this study, the self-efficacy of university students in responding to TOEFL questions is examined in relation to gender and participation in TOEFL courses. This study uses a descriptive design with a total sample of 200 university students from two large institutions who are majoring in both English and non-English [12].

III. Proposed Methodology

After reviewing data and determining the research aim and objectives, this paper examines the effects of characteristics such as education level, attendance, and student gender to examine students' scores in TOEFL iBT reading, listening, and writing using data mining approaches. For this study's techniques and data preparation procedures, methodologies are discussed below.

A. Dataset

The data for this study came from 473 students. Arabic is one of their first languages. 473 students in total took the TOEFL. The study enlisted the participation of 225 male and 248 female students (Table I).

TABLE I. Attributes of the Dataset

Attributes	**Details**
Gender	Male, Female
Education level	Faculty
Place	Cairo, Sheikh Zayed
Attendance	Number of Course Attendance

B. Data Preparation

All activities were taken from the raw data to create the final dataset (data that was entered into the design tool). The dataset's variables were prepared to generate the models needed in the next phase.

The students received a variety of English language skills, including a TOEFL preparation session, during the rigorous English language program. The TOEFL scores of the students were used as the research tool. At the end of the course, students take the TOEFL (paper-based test). Students were in class for five hours a day and were given TOEFL-related assignments. Listening, grammar/structure, and reading are the three skills that make up the TOEFL score. The TOEFL score ranges between 310 and 677. This study aims to determine the future effects of test education level, gender, place, and attendees on TOEFL iBT listening, reading, and writing results.

IV. Model and Algorithm

Fig. 1 depicts a framework for predicting student success. First, the data on student performance is fed into this system. This student data set has been preprocessed to eliminate noise and make the data set more consistent. The input data set is then subjected to various SPSS statistics analyses. Next, data analysis is carried out. Finally, different algorithms' categorization results are compared.

Likewise, gender is another factor that is usually studied, but there is a lack of good research to identify whether male and female language learners have significantly different TOEFL results. From a psychological standpoint, there are numerous variables related to gender [13].

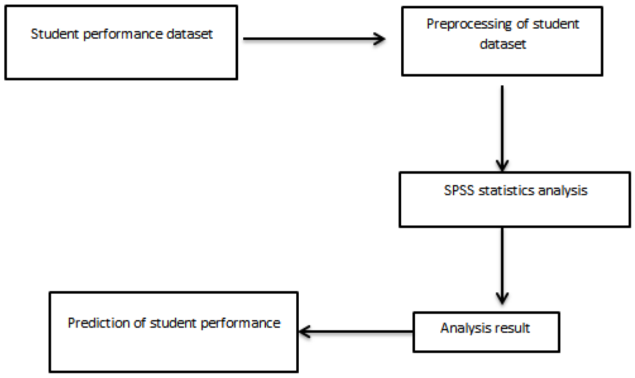

Student performance dataset
Preprocessing of student dataset
SPSS statistics analysis
Analysis result
Prediction of student performance

Fig. 1. A Framework for Student Performance Prediction.

In general, females are believed to be more successful in language learning than males. Therefore, many scholars in language acquisition studied how gender disparities can affect students' language learning proficiency. In other words, ten studies found that female students were superior to male students in reading comprehension. In contrast, five studies found that male students were superior [14,15] also undertook a quantitative study to see if there are any gender differences in TOEFL scores and found no significant differences. The Educational Testing Service (ETS), on the other hand, came to a different result.

According to the survey, female pupils are more advanced than male students [16]. Females, for example, outperformed males in writing and reading, though the difference was minor. On the other hand, Male students performed higher in terms of listening and comprehension, as well as vocabulary proficiency [17].

Additionally, a standardized English language assessment examination, such as the Test of English as a Foreign Language, is required at most English language colleges and universities (TOEFL). However, because there are few standardized evaluation measures for all candidates, English proficiency ratings are occasionally utilized for purposes other than evaluating the "abilities of non-native English speakers to use and understand English."

However, in the lack of standard ranking techniques for all candidates, the TOEFL score may be used as a stand-in for those criteria; the TOEFL score is occasionally employed as a predictor of how well a potential student will perform at a university. Even when the TOEFL is not used as the main measure of academic success, minimum TOEFL score requirements are frequently enforced.

Despite the fact that the underlying English-language communication abilities that TOEFL scores represent may be significantly more important to academic performance in specific areas, TOEFL score minimums for admission frequently do not vary among academic majors or fields of study. Requiring the same minimum TOEFL score whatever of a student's selected major may lead to the exclusion of otherwise talented students from academic programmers where academic achievement is not contingent on language competence [18]. For example, an increased TOEFL score is less correlated with academic success in college students than in other college students (possibly because English communication skills largely determine academic success in these areas). It may be reasonable to adopt the TOEFL score entry requirements. More lenient for engineering applicants, especially those who can show enough preparation through means other than a TOEFL score.

Despite the fact that course enrollment has tripled in the past 10 years, little is known about the impact of environment tests and attendance on learning. According to a recent study of college students, course attendance and the student place have an impact on the examination scores. Therefore, differences in student accomplishment between groups should be viewed with caution. This study adds to the body of knowledge by addressing a recurring problem of earlier research: determining the impact of various classroom test conditions on exam scores. The features of test environments are rarely described in previous research. This study compares test scores from students who took examinations off-campus with test scores from students who were called back to school for probationary exams within a semester [8].

V. Experiments and Results

The analysis of this paper was done using the statistical package for social sciences (SPSS V26) for both descriptive and inferential statistics. In this work, ANOVA was used as a statistical analysis method. Because this study examines the significance of group differences, it uses an ANOVA statistical model with a continuous dependent variable (TOEFL scores) and categorical independent factors.

Because this study tries to observe the interaction between gender differences, ANOVA is the most appropriate statistical procedure among the numerous varieties of ANOVA [19]. Pre and post TOEFL scores are within-subject factors, while male and female are between-subject variables. To address the first study question, a statistically significant mean difference between before and post TOEFL scores will be studied. After that, we'll look at the statistically significant mean difference between male and female TOEFL scores. The impacts will next be compared between the TOEFL scores of males and females.

Table II provides descriptive statistics for the selected variables, including the minimum (Min), maximum (Max), mean (M), standard deviation (SD), and coefficient of variation (CV) (M=48.36,SD=7.519,CV=1 5.55%),(M=47.24,SD= 7.972,CV=16.88%),(M=47.07,SD=8.354,CV=17.75%), (M=475.38,SD=70.869,CV=14.91 %) respectively.

Table II shows some descriptive statistics and bivariate correlations among the selected variables provided in this section.

TABLE II. Descriptive Statistics

	N	**Min**	**Max**	**Mean**	**SD**
Listening	473	24	68	48.36	7.519
Grammar	473	27	68	47.24	7.972
Reading	473	27	67	47.07	8.354
Total	473	300	653	475.38	70.869

TABLE III. MULTIPLE CORRELATIONS

		Listening	Grammar	Reading	Total
Listening	***Pearson Correlation***	1			
	P-value				
	N	473			
Grammar	***Pearson Correlation***	.647***	1		
	P-value	.000			
	N	473	473		
Reading	***Pearson Correlation***	.642***	.781***	1	
	P-value	.000	.000		
	N	473	473	473	
Total	***Pearson Correlation***	.847***	.911***	.909***	1
	P-value	.000	.000	.000	
	N	473	473	473	473

Table III displays the bivariate correlations between the study's primary variables; all of the correlations were statistically significant at 0.001. These correlations vary between.642 and.642, indicating that all variables in the study have substantial moderate to strong multiple correlations.

Furthermore, the results of the multiple regression were reported, and it can be noticed that all variables have significant positive effect on the total score since (P<0.001), as a result, the null hypothesis is rejected, and the alternative hypothesis is accepted in Table IV.

TABLE IV. REGRESSION COEFFICIENT

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
		B	***Std. Error***	***Beta***		
1	***(Constant)***	.334	1.234		.270	.787
	Listening	3.336	.032	.354	102.996	.000***
	Grammar	3.396	.038	.382	90.461	.000***
	Reading	3.257	.036	.384	91.499	.000***

*** P < 0.001

Table IV, the assumptions of this study were examined using multiple regression analysis in this part.

Table V, the F-test in ANOVA table confirms the significance of the model since (F=546827.6, P<0.001).

TABLE V. ANOVA TABLE

Model		**Sum of Squares**	**Df**	**Mean Square**	**F**	**Sig.**
1	***Regression***	2363572.748	3	787857.583	52762.172	.000^{b}
	Residual	7003.222	469	14.932		
	Total	2370575.970	472			

On the other side, the impact of demographic variables on the students' overall scores will be studied in this section. Finally, the normal distribution test was done utilizing Skewness and kurtosis tests to choose between parametric and nonparametric testing Table VI [20].

TABLE VI. TEST OF NORMALITY

	N	**Skewness**		**Kurtosis**	
		Statistic	***Std. Error***	***Statistic***	***Std. Error***
Total	473	.066	.112	-.756	.224

Table VI, the values of Skewness and kurtosis for the score were within the range of ±2, indicating that the total score was normally distributed, according to the normality statistics.

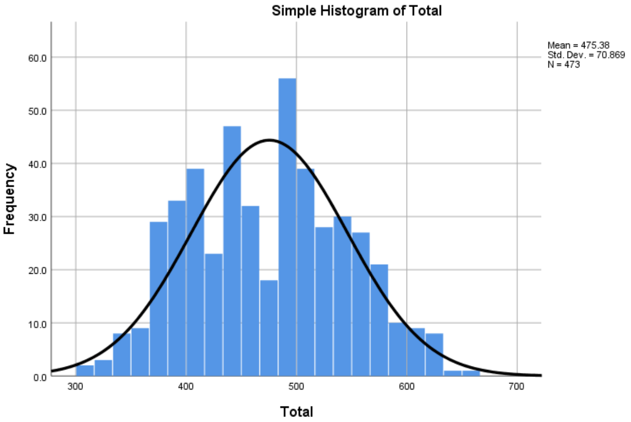


Fig. 2. Histogram and the Normal Curve of the Total Score.

Fig. 2 displys a normal distribution of data.

First hypothesis: there is a significant difference in total scores regarding the Gender of the students. The independent-samples t-test is the appropriate parametric test because Gender is a categorical variable with two independent categories.

Table VII, some descriptive statistics of the total score according to each category were given.

Fig. 3 can be concluded from that the average degree of females (487.49) was greater than that of males (462.04).

In addition, Levene's test for equality of variances was done and found that the variances were equal since $(F = .449, P > 0.05)$. The results of the independent-sample t-test show that there is a significant difference in total scores between males and females since P-value is less than 0.05 as $(t = -3.961, P < 0.001)$ Table VIII.

TABLE VII. DESCRIPTIVE STATISTICS OF THE TOTAL SCORES REGARDING THE GENDER

Gender	**N**	**Min**	**Max**	**Mean**	**Std. Deviation**
female	248	300	653	487.49	70.863
male	225	313	623	462.04	68.590
Total	473	300	653	475.38	70.869

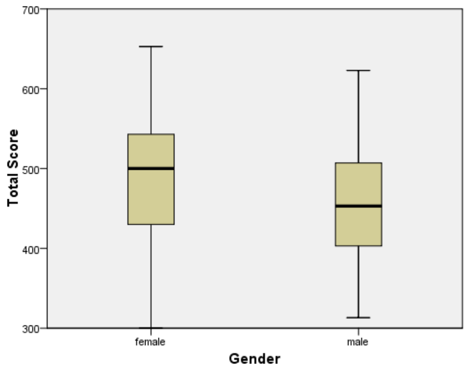


Fig. 3. Boxplot for the Total Scores of Students Regarding the Gender.

TABLE VIII. INDEPENDENT SAMPLES T-TEST

Levene's Test for Equality of Variances			t-test for Equality of Means					
	F	*P-value*	*t*	*df*	*Mean*	*Std. Error*	*95% CI of the Difference*	
							Lower	*Upper*
Equal variances assumed	.449	.503	-3.961	471	-25.452	6.426	-38.078	-12.826
Equal variances not assumed			-3.967	469	-25.452	6.415	-38.058	-12.845

***P < 0.001

In Table VIII, the results of the independent-sample t-test show that there is a significant difference in total scores between males and females since P-value is less than 0.05 as$(t = -3.961, P < 0.001)$.

Moreover, in the second hypothesis: there is a significant difference in total scores regarding the attendees of the students. Since the student's attendance is a categorical variable with more than two independent categories, the suitable parametric test is the analysis of variance (ANOVA) test.

In Table IX, some descriptive statistics of the total score according to each category were given.

Fig. 4 presented that the average scores of students attending for the first time (477.19) was greater than that of the second time (474.39), and the third time (473.07).

In Table X, the results of the ANOVA test show that there is no significant difference in total scores between the number of attendees since the P-value is greater than 0.05 as$(F = .151, P > 0.05)$.

TABLE IX. DESCRIPTIVE STATISTICS OF THE TOTAL SCORES REGARDING THE ATTENDANCE

Attendees	N	Minimum	Maximum	Mean	Std. Deviation
1	226	323	640	477.19	71.650
2	124	313	623	474.39	65.787
3	123	300	653	473.07	74.747
Total	473	300	653	475.38	70.869

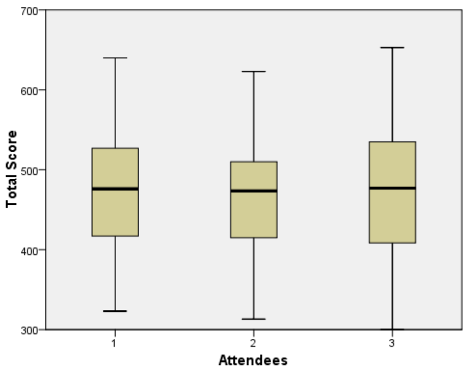


Fig. 4. Boxplot for the Total Scores of Students Regarding the Attendees.

TABLE X. ANOVA TABLE

	Sum of Squares	Df	Mean Square	F	P-value
Between Groups	1525.638	2	762.819	.151	.860
Within Groups	2369050.333	470	5040.533		
Total	2370575.970	472			

As well the third hypothesis: there is a significant difference in total score regarding the place of the test. Since the place of the test is categorical variable with two independent categories, so the suitable parametric test is the independent-samples t-test.

Table XI shows some descriptive statistics of the total score according to each category were given.

Fig. 5 displayed that the average scores of students attend in Sheikh Zayed (484.81) were greater than those attending Cairo (466.38).

TABLE XI. DESCRIPTIVE STATISTICS OF THE TOTAL SCORES REGARDING PLACE OF THE TEST

place	N	Min	Max	Mean	Std. Deviation
Cairo	242	300	640	466.38	71.684
Sheikh Zayed	231	350	653	484.81	68.905
Total	473	300	653	475.38	70.869

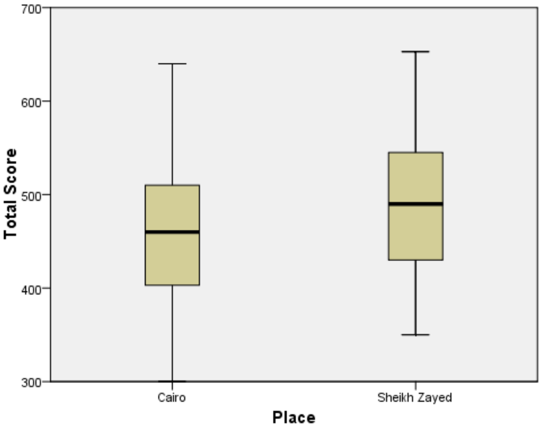


Fig. 5. Boxplot for the Total Scores of Students Regarding the Place of the Test.

TABLE XII. INDEPENDENT SAMPLES T-TEST

t-test for Equality of Means						
	t	df	Mean	Std. Error	95% CI of the Difference	
					Lower	*Upper*
Equal variances assumed	-2.848	471	-18.430	6.470	-31.144	-5.715
Equal variances not assumed	-2.851	471	-18.430	6.464	-31.132	-5.727

In Table XII, Levene's test for equality of variances reveals that the variances were equal since$(F = .475, P > 0.05)$. The results of the independent-sample t-test show that there is a significant difference in total scores between Cairo and Sheikh Zayed since P-value is less than 0.05 as$(t = -2.848, P < 0.01)$.

Subsequently, the fourth hypothesis shows a significant difference in total scores regarding the level of education**.** Since the level of education is a categorical variable with more than two independent categories, the suitable parametric test is the analysis of variance (ANOVA) test.

TABLE XIII. DESCRIPTIVE STATISTICS OF THE TOTAL SCORES REGARDING THE LEVEL OF EDUCATION

Faculty	N	Min	Max	Mean	SD
Academy of Arts	3	300	570	465.67	145.074
African Institute	15	313	653	446.73	87.928
Agriculture	10	380	510	452.70	47.579
Applied Arts	6	417	577	499.00	50.931
Arab Academy	4	430	517	475.00	40.406
Archaeology	5	350	600	476.60	89.960
Arts	17	393	553	483.47	50.222
Commerce	40	363	600	473.05	59.843
Computer and Information	10	383	580	501.30	58.317
Dar Al Uloom	6	400	620	464.33	85.141
Dentistry	17	450	610	542.35	43.566
Economics and Political Sciences	13	360	617	523.38	72.240
Education	5	383	607	465.40	84.145
Egyptian fellowship	2	500	513	506.50	9.192
Engineering	25	423	640	519.80	53.275
environment institute	1	417	417	417.00	.
Georgia	13	413	623	528.00	49.427
Grant	2	450	560	505.00	77.782
industrial education	1	410	410	410.00	.
Institute of Arabic Studies	1	450	450	450.00	.
Institute of Technical healthy	109	350	523	411.31	39.451
Kindergarten	2	420	497	458.50	54.447
laser institute	1	413	413	413.00	.
Law	10	347	547	438.10	75.253
MBA	5	500	560	514.00	26.077
media	30	390	620	496.63	69.042
Medicine	27	410	623	549.19	45.668
National Institute of Intellectual Property	1	503	503	503.00	.
natural medicine	3	390	413	403.33	11.930
Naval Academy	7	377	507	477.71	49.291
Nursing	6	367	503	431.17	45.512
Oncology Institute	1	573	573	573.00	.
Pharmacy	14	447	627	542.21	52.850
Physical Education	5	327	473	406.60	55.383
Postgraduate Education	12	410	553	488.00	42.988
Research Institute	11	327	573	453.91	61.119
Sadat Academy	4	450	500	472.50	26.300
Sciences	16	450	597	520.81	39.507
Social Service	1	453	453	453.00	.
Statistics Institute	6	387	563	517.33	67.666
Tourism and Hotels	2	480	500	490.00	14.142
urban planning	2	557	587	572.00	21.213
veterinary medicine	2	460	610	535.00	106.066
Total	473	300	653	475.38	70.869

Table XIII shows some descriptive statistics of the total score according to each category were given.

Fig. 6 concluded that students' average scores were different across the level of education.

Table XIV shows the results of the ANOVA test show that there is a significant difference in total scores across the level of education since the P-value is less than 0.05 as$(F = 8.407, P < 0.001)$.

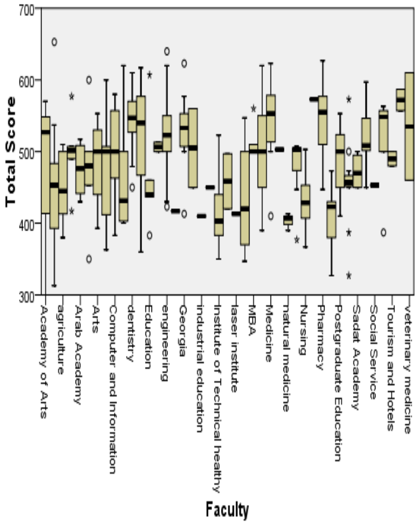


Fig. 6. Boxplot for Students' Total Scores Regarding the Level of Education.

TABLE XIV. ANOVA TABLE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1068908.842	42	25450.211	8.407	.000
Within Groups	1301667.129	430	3027.133		
Total	2370575.970	472			

Finally, the fifth hypothesis: there is a significant difference in TOEFL parts (Listening, Grammar, and Reading) regarding the Gender.

Table XV shows some descriptive statistics of the TOFEL parts according to each category were given.

TABLE XV. DESCRIPTIVE STATISTICS OF THE TOEFL PARTS REGARDING THE GENDER

Gender		Listening	Grammar	Reading
female	***N***	248	248	248
	Minimum	24	27	28
	Maximum	68	68	67
	Mean	49.36	48.58	48.41
	Std. Deviation	8.060	7.869	7.932
male	***N***	225	225	225
	Minimum	32	27	27
	Maximum	68	67	67
	Mean	47.26	45.76	45.59
	Std. Deviation	6.720	7.838	8.572
Total	***N***	473	473	473
	Minimum	24	27	27
	Maximum	68	68	67
	Mean	48.36	47.24	47.07
	Std. Deviation	7.519	7.972	8.354

Fig. 7, can be concluded that for Listening, the average degree of females (49.36) was greater than that of males (47.26), for Grammar, the average degree of females (48.58) was greater than that of males (45.76), and for Reading the average degree of females (48.41) was greater than that of males (45.59).

Then, Levene's test for equality of variances was conducted. It can be noticed that for listening, we have unequal variances since ($F = 7.566, P < 0.01$) but for Grammar, we have equal variances since($F = .007, P > 0.05$)and the same for Grammar. We have equal variances since($F = 1.870, P > 0.05$). The results of the independent-sample t-test show that there is a significant difference in listening scores between males and females since P-value is less than 0.05 as($t = -3.082, P < 0.01$). Moreover, there is a significant difference in grammar scores between males and females since P-value is less than 0.05 as($t = -3.900, P < 0.001$). Finally, there is a significant difference in reading scores between males and females since P-value is less than 0.05 as($t = -3.716, P < 0.001$) Tables XVI and XVII.

In Tables XVI and XVII, since the Gender of the students is categorical variable with two independent categories; the suitable parametric test is the independent-samples t-test.

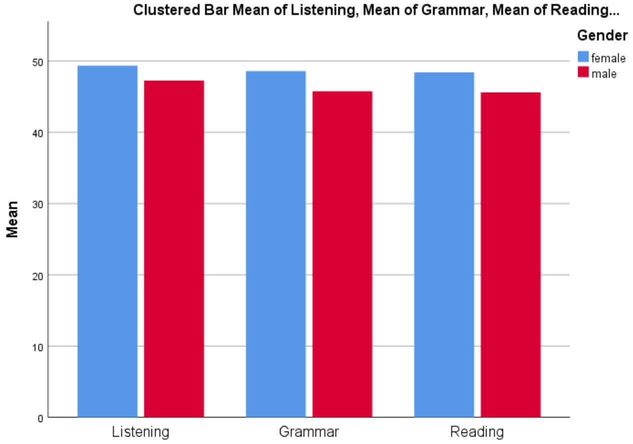


Fig. 7. Clustered Bar Chart for the TOEFL Parts Scores Regarding the Gender.

TABLE XVI. INDEPENDENT SAMPLES T-TEST

	Levene's Test for Equality of Variances		
		F	**P-value**
Listening	***Equal variances assumed***	7.566	.006
	Equal variances not assumed		
Grammar	***Equal variances assumed***	.007	.932
	Equal variances not assumed		
Reading	***Equal variances assumed***	1.870	.172
	Equal variances not assumed		

**P < 0.01

TABLE XVII. INDEPENDENT SAMPLES T-TEST

	t-test for Equality of Means							
		t	***df***	***P-value***	***Mean***	***Std. Error***	***95% CI of the Difference***	
							Lower	***Upper***
Listening	***Equal variances assumed***	-3.056	471	.002	-2.097	.686	-3.445	-.748
	Equal variances not assumed	-3.082	468	.002	-2.097	.680	-3.433	-.760
Grammar	***Equal variances assumed***	-3.900	471	.000	-2.820	.723	-4.241	-1.399
	Equal variances not assumed	-3.901	467	.000	-2.820	.723	-4.241	-1.399
Reading	***Equal variances assumed***	-3.716	471	.000	-2.820	.759	-4.311	-1.329
	Equal variances not assumed	-3.702	457	.000	-2.820	.762	-4.317	-1.323

**P < 0.01

VI. Conclusion

This study looked at the TOEFL results of 473 students based on how much time they spend studying, their educational level, gender, course attendance, and place. AS EXPECTED, the TOEFL scores improved from pre- to post-test, and the change was statistically significant. In this survey, there was significant difference by educational level, gender, attendance, and place difference. Furthermore, there was a relationship between male and female students' before and post TOEFL scores. As a result, the study's findings offer students with useful information. Furthermore, TOEFL educators can propose that the more time a student devotes to learning, the higher their TOEFL score will be. This also aids programmer makers in class design by giving them a sense of what students (who are prepared for the TOEFL) could expect. Because many students are applying to universities each year, generalizing TOEFL scores to the general population is insufficient.

References

[1] BACHMAN, Lyle F., et al. Fundamental considerations in language testing. Oxford university press, 1990.

[2] WEIR, Cyril J. Language testing, and validation. Hampshire: Palgrave McMillan, 2005.

[3] Choi, I.-C, “The impact of EFL testing on EFL education in Korea. Language Testing”, Inn-Chull Choi, vol.25, No.1, pp.39–62, January 2008.

[4] Messick, S , “Validity and washback in language testing. Language Testing”, Samuel Messick, vol.13, No.3, pp. 241-256, November 1996

[5] Kunnan, A. J, “Test fairness, test bias, and DIF. Language Assessment Quarterly”, vol.4, No.2, pp. 109–112, Dec 2007.

[6] Llosa, Lorena, and Margaret E. Malone. "Student and instructor perceptions of writing tasks and performance on TOEFL iBT versus university writing courses." Assessing Writing, vol.34, pp. 88-99, October 2017.

[7] BACHMAN, Lyle F., et al. Language testing in practice: Designing and developing useful language tests. Oxford University Press, 1996.

[8] IW Wait, JW Gressel, “Relationship between TOEFL score and academic success for international engineering students”, Journal of Engineering Education, vol.98, No.4, pp. 389-398, October 2009.

[9] Saeun, L. E. E. "Improvement of pre-and post-tests and gender differences on TOEFL scores." Bulletin of Miyazaki Municipal University Faculty of Humanities, vol.25, No.1, pp. 193-204, 2018.

[10] Neumann, Heike, Nina Padden, and Kim McDonough. "Beyond English language proficiency scores: Understanding the academic performance of international undergraduate students during the first year of study." Higher Education Research & Development, vol.38, No.2, pp.324-338, Sep 2019.

[11] Destiyanti, Cahya, Muhammad Amin, and Lalu Jaswadi Putera. "Gender-Based Analysis of Students’ Ability in Answering Factual and Vocabulary-in-Context Questions of the TOEFL-Like Reading Comprehension Test." PALAPA vol.9, No.1, pp.1-17, 2021.

[12] Yoestara, Marisa, and Zaiyana Putri. "Gender and language course participation differences in the university students’ self-efficacy of TOEFL." Journal of Physics: Conference Series. Vol. 123,No.1, October2019.

[13] MACCOBY, Eleanor E.; JACKLIN, Carol Nagy. The psychology of sex differences. Stanford University Press, 1978.

[14] Hyde, Janet S., and Marcia C. Linn. "Gender differences in verbal ability: a meta-analysis." Psychological bulletin, vol.104, No.1, pp.53-69, Jul 1988.

[15] Lin, J., & Wu, F. Differential Performance by Gender in Foreign Language Testing, 2004.

[16] Cole, N. S. The ETS gender study: how females and males perform in educational setting. Princeton, NJ: Educational Testing Service, 1997.

[17] Boyle, J. P. “Sex Differences in Listening Vocabulary. Language Learning”,vol.37,No.2,pp.273-284,June 1987.

[18] Simner, M.L.” Use of the TOEFL as a standard for university admission: A position statement by the Canadian Psychological Association”. European Journal of Psychological Assessment, vol.14, No.3, pp. 261–65, 1998.

[19] Lomax, R. G. & Hahs-Vaughn . An introduction to statistical concepts (3rd Ed). New York, NY: Routledge, 2012.

[20] Gravetter, F., & Wallnau, L. Essentials of statistics for the behavioral sciences (8t h ed.). Belmont, CA: Wadsworth, 2014.

Average Delay-based early Congestion Detection in Named Data of Health Things

Asmaa EL-BAKKOUCHI[1]*, Mohammed EL GHAZI[2]
Anas BOUAYAD[3], Mohammed FATTAH[4], Moulhime EL BEKKALI[5]
Artificial Intelligence, Data Sciences and Emerging Systems Laboratory
Sidi Mohamed Ben Abdellah University, Fez, Morocco [1,2,3,5]
IMAGE Laboratory, Moulay Ismail University, Meknes, Morocco[4]

***Abstract*—The Internet of Health Things (IoHT) is receiving more attention from researchers because of its wide use in the healthcare field. IoHT refers to medical devices whose main purpose is to transmit health data in a secure and lossless manner between them and healthcare personnel. However, in a medical emergency, sensors transmit vital patient data simultaneously and frequently, increasing the risk of congestion and packet loss. This problem is highly undesirable in an IoHT system, leading to undesirable results. To address this issue, a new approach based on Named Data Networking (NDN) (which is considered as the most appropriate internet architecture for IoT systems) is proposed to control congestion in IoHT systems. The proposed approach, Average delay-based early congestion Detection (ADCD), detects and controls congestion at consumer nodes by calculating the average queuing delay based on the one-way delay similar to that proposed in Sync-TCP. Then according to the calculated value, ADCD divides the network into three states: no-congested state, less congested state, and heavily congested state. The adjustment of the congestion window size is done according to the state of the network. ADCD was implemented in ndnSIM and compared to the Interest Control Protocol ICP. The results show that ADCD maximizes bandwidth utilization compared to ICP and maintains a reasonable delay.**

Keywords*—Named data networking; internet of health things; congestion control; congestion detection

I. Introduction

The Internet of Things (IoT) is a collection of intelligent devices that can communicate, interact and exchange information with each other [1]. It interconnects billions of small devices to deliver information at any time and across the world [2]. IoT has been primarily applied in several domains like healthcare, smart home, smart cities, industries, transportation and logistics, etc. In healthcare, IoHT has been proposed as an extended version of IoT to connect people and smart objects in the medical field [3]. It refers to medical devices, healthcare sensors and intelligent biomedical applications that enable the exchange and treatment of various types of healthcare data with each other and with healthcare people [3], whose main objective is to transmit health data securely and without loss. This data can be patient monitoring data, patient analyses, consultations or even sensitive data such as heart rate and breathing.

In the case of a medical emergency or patient vital signs monitoring, sensors implanted on patients detect and transmit vital patient data simultaneously and frequently, which increases the risk of congestion and consequently packet loss. In IoHT, congestion is highly undesirable and can lead to undesirable outcomes such as patient death. However, ensuring that packets arrive at their destination in time guarantees patients' safety and survival [4]. Due to the importance and sensitivity of the transmitted data, congestion should be avoided as much as possible and controlled if this is not possible.

However, the IP protocol stack on which the IoT is based was designed for a different purpose and cannot handle the important challenges caused by the heterogeneity of the devices and the importance of the traffic generated, highlighting the limitations of IP. However, the current IP solutions are working hard to support IoT systems, but the gaps in IP are still hard to hide. Along with efforts to adapt IP to the IoT and other content-based architectures, Information Centric Networking (ICN) [5] promises to support IoT systems and emerging internet applications natively. It is a switch from a host-centric communication model to a content-centric system, caching at intermediate nodes, and the use of multiple paths and sources. With its features, ICN has the potential to be a reliable framework for IoT by connecting billions of heterogeneous constrained objects. Indeed, ICN provides easy access to data and reduces data recovery time and the load on data producers. NDN [6] is considered as the appropriate ICN architecture for IoT systems among several ICN architectures.

NDN is a new internet architecture based on the content of data rather than its IP address. NDN defines three roles: Consumer, Router and Producer and offers two types of packets; the interest packet, containing the required content name and the data packet, containing the required content. There are three elements to every NDN node as shown in Fig. 1; Content Store (CS), Pending Interest Table (PIT) and Forwarding Interest Base (FIB). The CS stores copies of the content that passes through it so that future demands for the same content can be satisfied. The PIT is a table that preserves the records of incoming and outgoing interest and data packets. The FIB transfers interest and data packets between nodes through routing protocols [7]. The data transfer process between NDN nodes is as follows, the consumer asks for content by sending an interest packet.

*Corresponding Author.

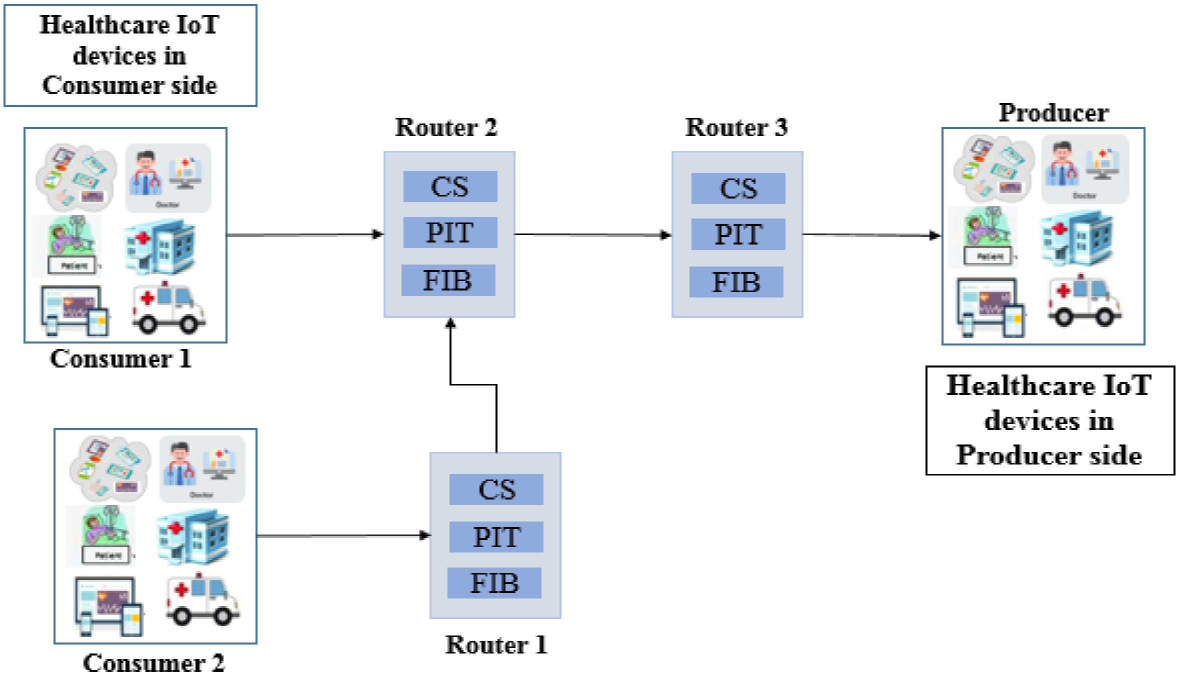

Healthcare IoT devices in Consumer side
Consumer 1
Router 2
CS
PIT
FIB
Router 3
CS
PIT
FIB
Producer
Healthcare IoT devices in Producer side
Consumer 2
CS
PIT
FIB
Router 1

Fig. 1. NDN Information Distribution in Case of IoHT Environment.

The routers use FIBs to forward this packet to the content producer and create a PIT entry list on each router to define the reverse path. Then, the producer transfers the related content via the reverse path to the consumer, and the CS saves the content that traverses it for future utilization [8].

An example of the data transfer process between NDN nodes is shown in Fig. 1. Suppose that consumer one starts this process and sends an interest packet containing the requested name. This interest packet is transferred to the producer using the FIB services. The requested content is put into a data packet and is stored in the CS of router three and router two for future utilization. Subsequently, if consumer two requires the same content, its interest packet is locally satisfied from the CS of router two without the need to forward this packet to the original content producer. In the case of IoHT, the consumer can be a doctor, a patient, a nurse, a medical application, a hospital or any medical or IoT device that desires to obtain data/information concerning the health field. The producer can also be a doctor, a patient, a nurse, a medical application, a hospital or any medical or IoT device that has data/information concerning the health field. The router's mission is to route the data/information concerning the health field between the consumer and the producer.

To address the problem of congestion in healthcare field, and as NDN congestion control mechanisms that are based on the estimation of RTTs as the main indication of congestion are not reliable in NDN-IoHT because the RTT value changes frequently due to caching or multipath. This paper proposes a new approach based on Sync-TCP [9] (which has been proposed to control congestion based on measurements of the One-way Transit Time (OTT) between senders and receivers of packets) to control congestion in NDN-IoHT. OTTs are more accurate in reflecting queue delay resulting from network congestion than RTTs.

In our approach, congestion is detected by calculating the average queuing delay based on the one-way delay similar to that proposed in Sync-TCP [9]. Then according to the calculated value, the network is divided into three states; no-congested state, less-congested state and heavily congested state. The adjustment of the congestion window size is made according to the state of the network. Increased in case of a no-congested network and decreased in a less-congested and heavily congested network. The choice of congestion control at the consumer nodes comes from the fact that the method used for congestion detection is based on variations of OTTs which is more reliable at the consumer node than at the router node because consumers are able to estimate the one-way delay of the data packet using the transmission time available in its header and the reception time of this packet. The rest of the paper is organized as follows: Section II presents background and related work, Section III describes the proposed method, Section IV presents results and discussion and Section V concludes the article.

II. Background and Related Work

A. NDN for IoT

ICN has been proposed as a promising future internet architecture to fill the gaps in the CoAP/RPL/6LowPan/802.15.4 protocol stack on which the IoT system is based and improve its deployment and data distribution [10]. Among the ICN architectures proposed in the literature, NDN is considered as the most appropriate ICN architecture for IoT systems. In effect, the characteristics of NDN, namely hierarchical naming of unique content that is independent of its location, caching in intermediate routers, multipath, multisource, name-based routing, support for user mobility, the use of encryption for better access control, make it a very suitable platform for IoT system traffic and applications.

Several works have been done on this topic to show that NDN is the most suitable and promising architecture for IoT. In [11], the authors suggested that NDN can meet IoT requirements and is the most suitable architecture for IoT scenarios. In [12], the authors showed that the semantics of NDN meets the requirements of IoT applications and its main challenges. They also showed that the communication model used by NDN "based on the content name" allows IoT networks an easy deployment and configuration. In [13], the authors proposed a comparative study of ICN architectures in an IoT context and concluded that NDN is an architecture most suitable for IoT systems. In [2], the authors proposed an NDN integration in IoT devices and then evaluated this proposal in a Smart Farming application scenario to prove that NDN is an architecture most suitable for IoT systems. In [14], the authors described the advantages of the NDN architecture compared to the current IP internet architecture for IoT systems and then explained how NDN can be included in an IoT architecture. In [15], the authors discussed the main characteristics of NDN in IoT, then they proposed an IoT architecture via NDN named IoT-NDN for various IoT domains. Some studies have discussed the requirements of IoT and how ICN architectures support them without examining which of these architectures is most appropriate for IoT.

B. *Related Work*

In NDN architecture, the problem of network congestion is the subject of much active research focused on minimizing transit delays and reducing packet loss caused by the transfer capability of routers. In NDN architecture, data packets are the leading cause of congestion because they are much larger than interest packets. Therefore, to control congestion, many research works propose controlling the sending rate of interest packets to limit the returning rate of data packets. Several congestion control mechanisms have been proposed in the literature, classified as receiver-based control, Hop-by-hop control, and Hybrid control [16]. In this paper, we present some NDN congestion control works. Among the first interest control protocols proposed for NDN is the Interest Control Protocol (ICP) [17], which detects congestion at the consumer node by measuring delay and timer expirations. The window size adjustment is made at the receiver level using the AIMD (Additive-Increase Multiplicative-Decrease) mechanism. If RTO (Retransmission Time-Outs) is triggered, the consumer decreases its congestion window by MD (Multiplicative Decrease). Otherwise, it increases its congestion window by AI (Additive Increase). The authors of [18] proposed DCP "Delay-based Congestion Control Protocol" based on the window and the receiver. DCP detects congestion based on the value of the queue delay. If this delay is below a given threshold, DCP considers that the link is not congested. Otherwise, DCP considers the link is congested. The calculation of queuing delay is done by measuring the delay returned by the producer or intermediate nodes along the path of the transmitted data packets. DCP uses a linear controller to adjust the congestion window. In [19], the authors proposed a hop-by-hop congestion control mechanism (HCCM) based on explicit notification of interest packet rates. This mechanism detects congestion by calculating the queue length of the interest packets of the output interface. According to the value found, each router between the congested node and the consumer can adjust the sending rate of interest packets. The authors used two levels in the queue, qmax and qmin, representing the maximum and minimum queue occupancy thresholds respectively. Once the queue reaches one of the levels (qmax or qmin), the router sends a notification to the downstream node to inform it of the congestion state and the regulation of the sending rate of interest packets. In [7], the authors proposed EC-Elastic, an Explicit Congestion Control Mechanism that detects congestion at the routers by measuring the sojourn time of packets in the queue using CoDel-AQM algorithm and then, according to this value, the router marks the concerned data packets to inform the consumer nodes to reduce their sending rate of interest packets. At the consumer node, if the phase is a slow start, the congestion window is increased by one and decreased by $\beta_{1,}$ and if the phase is the congestion avoidance, the congestion window is increased by $\frac{WWF}{cwnd}$ using the Window-correlated Weighting Function WWF of Elastic-TCP [20] and decreased by β2. In [21], the authors proposed the MPCC Multipath Congestion Control mechanism, which is based on two principles: Multipath discovery that tags each sub-path with a path tag in the forwarding process and then based on these tags, a tag-aware forwarding strategy has been proposed to discover and manage sub-paths. For multipath congestion control, the authors proposed a Multipath Window Adaptation Control (MWAC) scheme to control the congestion window. In [22], the authors proposed DPCCP Delay-based Path-specified Congestion Control Protocol based on three modules, namely: The congestion estimation module, which aims to measure the number of backlogged packets for every sub-flow using RTT and baseRTT, where the number of backlogged packets measured for a sub-flow is the product of the sub-flow queuing delay and the sub-flow rate. The fairness control module which is used to calculate the target number of backlogged packets for each sub-flow to equalize the aggregate queuing delay [23] and the flow control module, which aims to adjust the rate based on the queuing delay and the target number of packets in the queue using the Adaptive Additive Increase Additive Decrease (A-AIAD) algorithm.

Different from prior works, this paper proposes a new approach that controls congestion by calculating the average queuing delay based on the one way delay similar to that proposed in Sync-TCP [9]. Then according to the calculated value, the network is divided into three states; no-congested state, less-congested state and heavily congested state for efficient and lossless deployment in an IoHT environment. The adjustment of the congestion window size is made according to the state of the network. Increased in case of a no-congested network and decreased in a less-congested and heavily congested network.

III. The Proposed Method

A. *Motivation*

The NDN paradigm has several features, such as caching in intermediate routers that serve future requests for the same content without going through the content producer. This feature reduces the content retrieval time and the charge on the content producer. The use of multiple paths and multisource to avoid congestion of one path over another and to divert requests in the event of a congested path. However, the use of RTT (Round-Trip Time) estimates as a primary indication of congestion are unreliable in NDN-IoHT because the value of RTT frequently changes due to caching and multipath causing problems for congestion control mechanisms that are affected by RTT variation particularly in the event of large RTTs which are treated as losses despite having no packet losses, thus halving the rate of sending packets of interest and consequently more time to fully utilize the bandwidth.

An explicative example of the inefficiency of this congestion detection method was discussed in [18]. In this article, the authors shows that the increase in RTT along the path can be caused by the PIT congestion of one of the routers and not by the data path congestion. In addition, when using caching, the content retrieved from the cache has a shorter RTT. This situation can lead to an erroneous increase in the congestion window size because a short RTT value is perceived as a sign of bandwidth availability, which can cause packet loss [18].

Our approach is motivated by these problems of considering RTT as an indication of congestion, which is not tolerated in an IoHT environment because the transmitted data are important and critical and the recovery time plays a crucial role in this domain. In an NDN-IoHT environment, a congestion control mechanism should achieve the following goals: Avoid congestion if possible and, in cases where congestion cannot be avoided, control it; Ensure delivery of healthy data transmitted in the network and minimize packet loss; Utilize available bandwidth and maintain low packet delivery delay.

B. *Congestion Detection*

The congestion detection method of ADCD is inspired by the Sync-TCP algorithm proposed in [9]. It detects congestion based on One-way Transit Time (OTT) measurements. OTTs are more accurate in reflecting queue delay resulting from network congestion than RTTs. In NDN, each node (router or producer) adds the time of its transmission in the header of each data packet so that the consumer can estimate the one-way delays. When the consumer receives this packet, it can estimate the one-way delay of the data packet using the transmission time available in its header and the reception time of this packet by subtracting the reception time of the packet from the sending time of that packet as follows:

$$OTT = A_T - T_T \tag{1}$$

where, A_T refers to the data packet arrival time and T_T refers to the transmission time of this data packet. Similar to DCP [18], we are focused only on the measurement of the relative delay between data packets to know whether the content is served from a new data producer or not. ADCD uses changes in queuing delay of a path to detect congestion. For each data packet received, the consumer ADCD obtains a new estimate of the queuing delay and then calculates it average. The average queuing delay is based on the changes in the current queuing delay measure as follows:

$$AvgQDelay = 0.875 * AvgQDelay + 0.125 * CurrQDelay \tag{2}$$

Where CurrQDelay is the current queuing delay, it is estimated by taking the minimum observed OTT and subtracting it from the current delay as follows:

$$CurrQDelay = \text{Current\_delay} - OTT_{min} \tag{3}$$

Where Current_delay is the one-way delay incurred by data packets and OTT_{min} is the minimum one-way delay. The current queuing delay CurrQDelay is calculated for each received data packet once the minimum OTT and current_delay are updated, as shown in Algorithm 1. Once AvgQDelay is calculated, ADCD situates it in one of the network states.

Algorithm 1 AvgQDelay Estimation Algorithm

```
Function ONDATA(DataPacket)
    Current_delay ← CurrentTime
    $OTT_{min}$ ← min ($OTT_{min}$, Current_delay)
    $OTT_{max}$ ← max ($OTT_{max}$, Current_delay)
    $OTT_{mid}$ ← ($OTT_{min}$ + $OTT_{max}$) /2
    CurrQDelay ← Current_delay - $OTT_{min}$
    AvgQDelay ← 0.875 *AvgQDelay + 0.125 * CurrQDelay
end function
```

ADCD considers three network congestion states; no-congested state, less-congested state and heavily congested state. ADCD determines these three states based on the changes in the received OTTs and adjusts the congestion window size cwnd according to the degree of congestion in the network. The role is as follows:

$AvgQDelay \le OTT_{mid}$ no-congested state

$OTT_{mid} < AvgQDelay < OTT_{max}$ less-congested state

$AvgQDelay \le OTT_{max}$ heavily congested state

If $AvgQDelay \le OTT_{mid}$, ADCD considers the network non-congested. In this status, ADCD has a low queuing delay, a sign of bandwidth availability. In this case, the ADCD consumer increases its congestion window according to the network phase (Slow Start or Congestion Avoidance) to utilize the available bandwidth.

If $OTT_{mid} < AvgQDelay < OTT_{max}$, ADCD considers the network low congested. As the risk of packet loss is low, the congestion window is decreased by the factor $\beta_{1.}$

If $AvgQDelay > OTT_{max}$, ADCD considers the network heavily congested and decreases its congestion window by the factor β2.

Where, OTT_{max} is the maximum one-way delay and OTT_{mid} is the mid-point between OTT_{max} and OTT_{min}. The values of OTT_{max}, OTT_{min} and OTT_{mid} are updated whenever a data packet is received by the consumer according to the following formulas:

$$OTT_{min} = \min(\text{Current\_delay}, OTT_{min}) \tag{4}$$

$$OTT_{max} = \max(\text{Current\_delay}, OTT_{max}) \tag{5}$$

$$OTT_{mid} == (OTT_{min} + OTT_{max})/2 \tag{6}$$

C. Adjustment of the Congestion Window

When the consumer calculates AvgQDelay and locates the status of the network, it proceeds to the adjustment of the congestion window size. ADCD controls the congestion window size in two phases Slow Start and Congestion Avoidance. ADCD adopts the Additive Increase mechanism in Slow Start phase and increases its congestion window by one while cwnd <= ssthresh, where ssthresh is a predefined threshold. In the congestion avoidance phase, which corresponds to cwnd > ssthresh, ADCD adopts the Window-correlated Weighting Function WWF of [20] and increases its congestion window by $\frac{WWF}{cwnd}$. The objective of using this function is to improve bandwidth utilization. Its formula is as follows:

$$WWF = \sqrt{\frac{\text{RTTmax}}{\text{RTT current}} * \text{cwnd}} \qquad (7)$$

where $\text{RTT}_{\text{current}}$ is the current RTT, RTT_{max} is the maximum RTT and cwnd is the last congestion window size.

ADCD starts the communication with the Slow Start phase and increases its congestion window by one by sending an interest packet to the network. Once the corresponding data packet is received, the consumer extracts the sending time and calculates AvgQDelay according to equation 2. If AvgQDelay is less than or equal to OTT_{mid}, the network is considered not congested. In this case, the consumer continues to increase the congestion window exponentially to use the available bandwidth fully. Otherwise, the network is considered less congested if AvgQDelay is between OTTmid and OTTmax. In this case, the consumer decreases its congestion window by the factor β_1= 0.8. Otherwise, if AvgQDelay exceeds OTT_{max}, the network is considered heavily congested, and the consumer decreases its congestion window by the factor β_2 = 0.5 and enters into the congestion avoidance phase to slowly increase its congestion window. In this phase, the congestion window increases by using the WFF function. If the consumer receives a NACK packet or TimeOut, The congestion window is divided by two using a multiplicative decrease. Algorithm 2 summarizes the congestion window adjustment steps at the consumer node.

Algorithm 2 CWND Adjustment Algorithm

```
On data reception do
    if slow start then
        cwnd ← cwnd + 1
    else
        cwnd ← cwnd + $\frac{WWF}{cwnd}$
    end if
    if NACK or TimeOut received then
        cwnd ← cwnd / 2
    end if
```

IV. Results and Discussion

In this section, we present the experiment's parameters and then evaluate the performance of the proposed mechanism ADCD in different scenarios using ndnSIM [24], an NS3-based simulator that has been proposed for NDN networks. We compare the performance of ADCD to that of ICP [17] in terms of throughput, delay, and packet loss rate.

A. Simulation Parameters

The first scenario is shown in Fig. 2. It represents the ideal case of a non-congested network for evaluating the performance of ADCD to demonstrate the high bandwidth utilization that can be achieved under the optimal conditions. It contains a consumer, a router, and a content producer. The bandwidth of the consumer-router path is 60 Mbps with a delay of 10ms, and the bandwidth of the router-producer path is 100 Mbps with a delay of 10ms. The second scenario is illustrated in Fig. 3. This scenario contains three consumers who request the same content, two routers, and two content producers. The three consumers have an equal bandwidth of 50Mbps with different delays, 10ms for consumer1, 5ms for consumer2, and 10ms for consumer3. The size of the transmitted data packet is 1024 bytes, and the simulation time was in the 30s.

B. Throughput

Throughput refers to the total number of packets successfully transmitted between the source and the destination every second. Fig. 4 compares the throughput between ADCD and ICP in the first scenario, while Fig. 5 compares ADCD and ICP in the second scenario. In both figures, time in seconds is defined on the x-axis, and throughput in Mbps is defined on the y-axis.

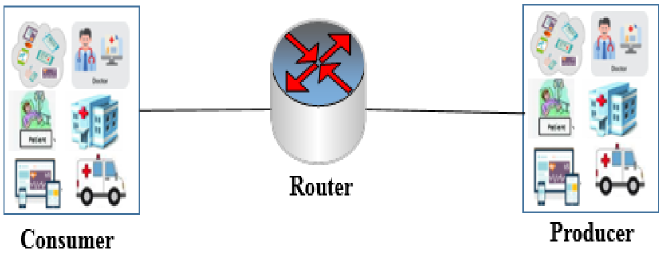


Fig. 2. Topology of the 1st Scenario.

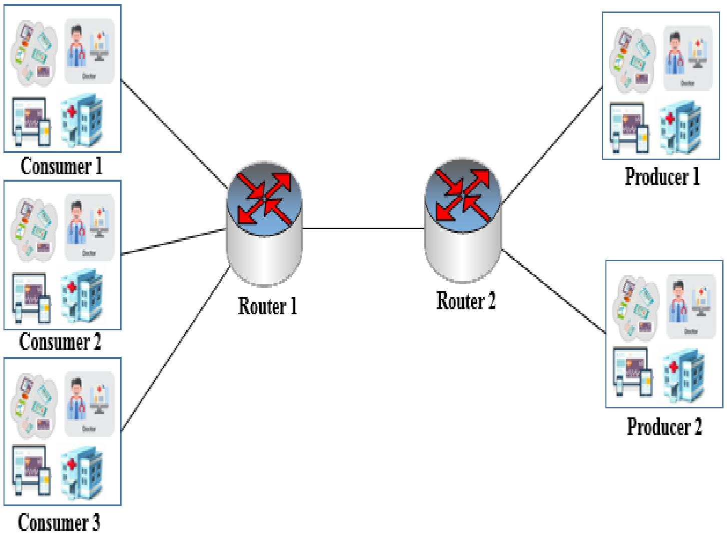


Fig. 3. Topology of 2nd Scenario.

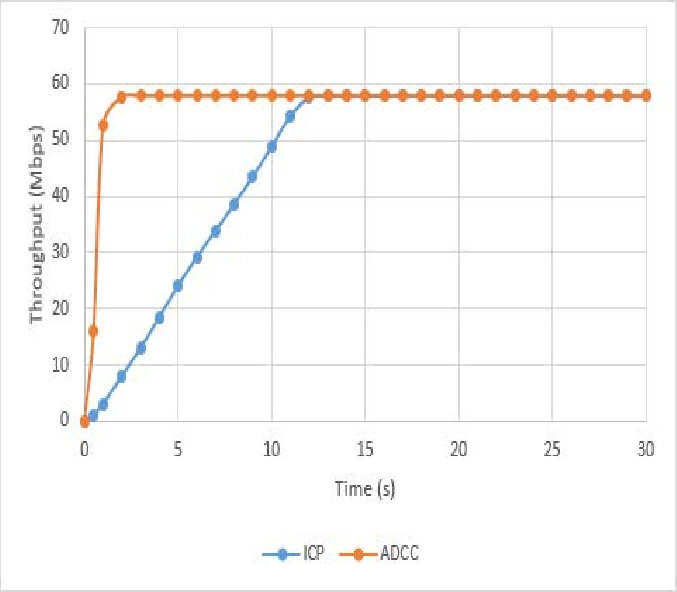


Fig. 4. Throughput of the 1st Scenario.

Fig. 4 shows the ability of the proposed mechanism to use the available bandwidth fully in the case of a non-congested network. This figure clearly shows that ADCD outperforms ICP in terms of throughput. The ADCD consumer could use an average throughput of 56.33 Mbps of the link capacity, while ICP used 45.63 Mbps. The increase in throughput with ADCD compared to ICP is due to the algorithm used by ADCD for congestion window adjustment, which rapidly increases its congestion window in the slow start phase, and then in the congestion avoidance phase, uses a window correlated weighting function (WWF) which aims to increase the use of the available bandwidth. However, ICP uses the AIMD algorithm to increase its congestion window, which rapidly increases the congestion window in the slow start phase, but in the congestion avoidance phase, it increases the congestion window by 1/cwnd, which increases the window slowly and consequently, a lower throughput than ADCD.

Similarly, in Fig. 5, which represents the throughput of the second scenario, we observe that the ADCD mechanism achieves better throughput than ICP. This is explained by the fact that ADCD divides the network into three states; no-congested state, less congested state, and heavily congested state. When it sees an increase in average queuing delay, it situates this value in one of the three states and reacts quickly by adjusting the size of the congestion window by β_1 or β_2. However, in ICP when a timeout is detected, it divides the congestion window by two, dividing the amount of data to be transmitted in the network by two, consequently decreasing the throughput.

These results demonstrate the capability of ADCD to manage the NDN network by transferring packets over the network bandwidth without causing network congestion or queue overflow. In the case of a congested network, ADCD responds quickly to the congestion problem while maintaining an optimal throughput.

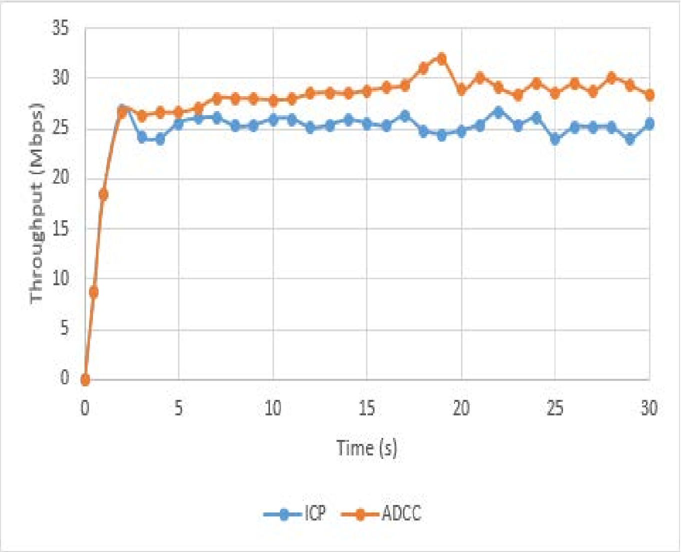


Fig. 5. Throughput of 2nd Scenario.

C. Delay Measurement

Delay is an essential factor for healthcare applications. It is the time taken for a packet to reach its destination from the source. Fig. 6 presents the delay measurements of the first scenario, Fig. 7 presents the delay measurements of the second scenario, and the average delay of scenarios 1 and 2 is presented in Table I.

In Fig. 6, we observe that ICP has a lower delay than ADCD, the consumer ICP has an average delay of 0.054s while the consumer ADCD has an average delay of 0.072s. In Fig. 7, we observe that ADCD has a lower delay than ICP. The consumer ADCD has an average delay of 0.26s while ICP is 0.42s. This is explained because ICP slowly increases its congestion window to transmit packets with lower throughput, fewer packets circulate in the network, and a lower transmission delay. However, the method used by ADCD to adjust the congestion window aims to increase the use of available bandwidth, thus allowing medical data to be transmitted with higher throughput and a reasonable transmission delay.

In the second scenario, the number of consumers has increased, so more requests for content are circulating in the network. ADCD handled this situation by calculating the average queuing delay for early congestion detection and reacting before overflowing the queue, consequently an optimal data transmission delay.

TABLE I. AVERAGE DELAY OF BOTH SCENARIOS

Mechanisms	**ADCD**	**ICP**
Scenario 1	0,072	0,054
Scenario 2	0,26	0,42

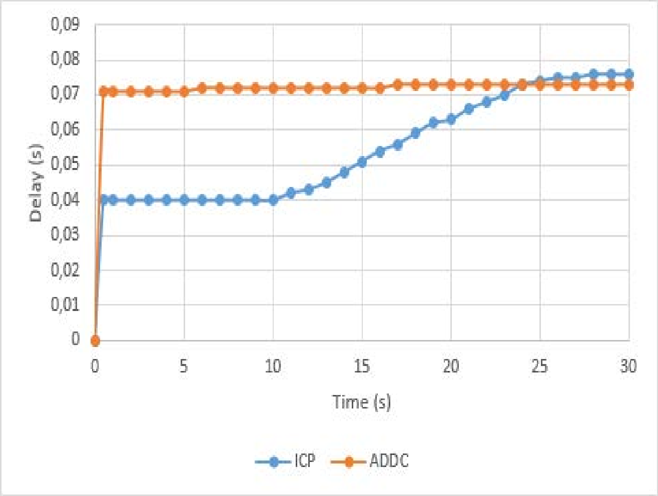


Fig. 6. Delay of the 1st Scenario.

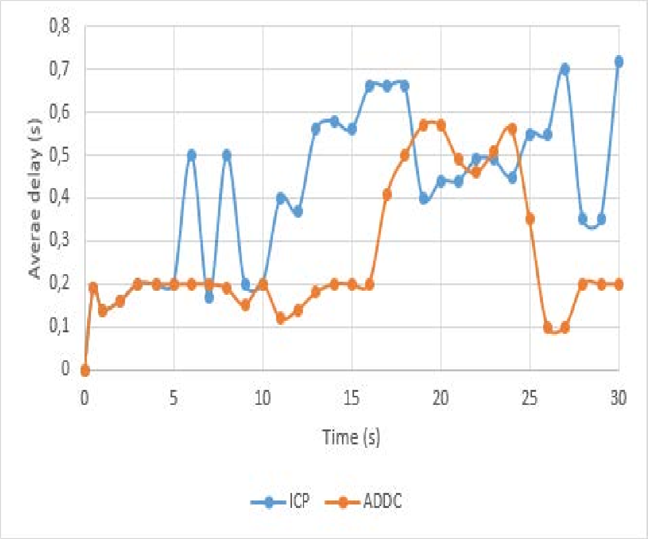


Fig. 7. Delay of 2nd Scenario.

D. Packet Loss Rate

The packet loss measurement is the difference between the number of packets transmitted and the number of packets received by the same node. It represents the number of packets abandoned per second. Congestion control mechanisms aim to reduce packet loss rate while increasing throughput and the use of available bandwidth. The simulation results of packet loss rate of the two scenarios are presented in Table II, where it is observed that the performance of ADCD and ICP are practically similar with a negligible packet loss rate.

TABLE II. PACKET LOSS RATE OF BOTH SCENARIOS

Mechanisms	ADCD	ICP
Scenario 1	0	0
Scenario 2	0,042	0,087

V. CONCLUSION

This paper proposes a new NDN-based approach, an Average delay-based early congestion Detection (ADCD), to control congestion in IoHT systems in which congestion is highly undesirable and can lead to undesirable results. In this approach, congestion is detected and controlled at consumer nodes by calculating the average queuing delay based on the one-way delay similar to that proposed in Sync-TCP, and then according to the calculated value, ADCD divides the network into three states: no-congested state, less congested state and highly congested state. The adjustment of the congestion window size is made according to the network state. The different simulations performed show the effectiveness of ADCD for early detection and control of congestion while improving network bandwidth utilization, maintaining optimal delay, and low packet loss rate.

In future work, we envisage extending the multipath analysis to independently managed paths to exploit the capability of NDN "multipath" to manage complex communication scenarios in IoHT systems.

REFERENCES

[1] K. Das, S. Zeadally, and D. He, “Taxonomy and analysis of security protocols for Internet of Things,” Futur. Gener. Comput. Syst., vol. 89, pp. 110–125, 2018.

[2] A. Abane, M. Daoui, S. Bouzefrane, S. Banerjee, and P. Muhlethaler, “A realistic deployment of named data networking in the internet of things,” J. Cyber Secur. Mobil., vol. 9, no. 1, 2020.

[3] Aroosa, S. S. Ullah, S. Hussain, R. Alroobaea, and I. Ali, “Securing NDN-Based Internet of Health Things through Cost-Effective Signcryption Scheme,” Wirel. Commun. Mob. Comput., vol. 2021, no. April, 2021.

[4] A. A. Rezaee, M. H. Yaghmaee, A. M. Rahmani, and A. H. Mohajerzadeh, “HOCA: Healthcare aware optimized congestion avoidance and control protocol for wireless sensor networks,” J. Netw. Comput. Appl., vol. 37, no. 1, pp. 216–228, 2014.

[5] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, “A survey of information-centric networking,” IEEE Commun. Mag., vol. 50, no. 7, pp. 26–36, 2012.

[6] P. Zhang, L., Afanasyev, A., Burke, J., Jacobson, V., claffy, k., Crowley, P. and B. C., Wang, L., Zhang, “Named data networking,” ACM SIGCOMM Comput. Commun. Rev., vol. 44, no. 3, pp. 66–73.

[7] A. EL-BAKKOUCHI, M. EL GHAZI, A. BOUAYAD, M. FATTAH, and M. EL BEKKALI, “EC-Elastic an Explicit Congestion Control Mechanism for Named Data Networking,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 11, pp. 594–603, 2021.

[8] A. El-bakkouchi, A. Bouayad, and M. ELBekkali, “A hop-by-hop Congestion Control Mechanisms in NDN Networks – A Survey,” 2019 7th Mediterr. Congr. Telecommun., pp. 1–4, 2019.

[9] M. C. Weigle, K. Jeffay, and F. D. Smith, “Delay-based early congestion detection and adaptation in TCP: Impact on web performance,” Comput. Commun., vol. 28, no. 8, pp. 837–850, 2005.

[10] B. Alahmri, S. Al-Ahmadi, and A. Belghith, “Efficient Pooling and Collaborative Cache Management for NDN/IoT Networks,” IEEE Access, vol. 9, pp. 43228–43240, 2021.

[11] M. Amadeo, C. Campolo, A. Iera, and A. Molinaro, “Named data networking for IoT: An architectural perspective,” EuCNC 2014 - Eur. Conf. Networks Commun., no. July 2015, 2014.

[12] W. Shang et al., “Named data networking of things (invited paper),” Proc. - 2016 IEEE 1st Int. Conf. Internet-of-Things Des. Implementation, IoTDI 2016, pp. 117–128, 2016.

[13] M. Meddeb, A. Dhraief, A. Belghith, T. Monteil, K. Drira, and S. Al-Ahmadi, “Named data networking: A promising architecture for the internet of things (IoT),” Int. J. Semant. Web Inf. Syst., vol. 14, no. 2, pp. 86–112, 2018.

[14] S. K. Datta and C. Bonnet, "Integrating Named Data Networking in Internet of Things architecture," 2016 IEEE Int. Conf. Consum. Electron. ICCE-TW 2016, 2016.

[15] M. A. Hail, "IoT-NDN: An IoT architecture via named data netwoking (NDN)," Proc. - 2019 IEEE Int. Conf. Ind. 4.0, Artif. Intell. Commun. Technol. IAICT 2019, no. July, pp. 74–80, 2019.

[16] Y. Ren, J. Li, S. Shi, L. Li, G. Wang, and B. Zhang, "Congestion control in named data networking – A survey," Comput. Commun., vol. 86, pp. 1–11, Jul. 2016.

[17] G. Carofiglio, M. Gallo, and L. Muscariello, "ICP: Design and evaluation of an Interest control protocol for content-centric networking," in 2012 Proceedings IEEE INFOCOM Workshops, 2012, pp. 304–309.

[18] A. A. Albalawi and J. J. Garcia-Luna-Aceves, "A Delay-Based Congestion-Control Protocol for Information-Centric Networks," 2019 Int. Conf. Comput. Netw. Commun. ICNC 2019, pp. 809–815, 2019.

[19] S. Mejri, H. Touati, N. Malouch, and F. Kamoun, "Hop-by-hop congestion control for named data networks," Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. AICCSA, vol. 2017-Octob, pp. 114–119, 2018.

[20] M. A. Alrshah, M. A. Al-Maqri, and M. Othman, "Elastic-TCP: Flexible Congestion Control Algorithm to Adapt for High-BDP Networks," IEEE Syst. J., pp. 1–11, 2019.

[21] F. Wu, W. Yang, M. Sun, J. Ren, and F. Lyu, "Multi-Path Selection and Congestion Control for NDN: An Online Learning Approach," IEEE Trans. Netw. Serv. Manag., vol. 18, no. 2, pp. 1977–1989, 2021.

[22] Y. Ye, B. Lee, R. Flynn, J. Xu, G. Fang, and Y. Qiao, "Delay-Based Network Utility Maximization Modelling for Congestion Control in Named Data Networking," IEEE/ACM Trans. Netw., pp. 1–14, 2021.

[23] Y. Cao, M. Xu, and X. Fu, "Delay-based Congestion Control for Multipath TCP," 20th IEEE Int. Conf. Netw. Protoc., pp. 1–10, 2012.

[24] S. Mastorakis, A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM 2 : An updated NDN simulator for NS-3," Dept. Comput. Sci., Univ. California, Los Angeles, Los Angeles, CA, USA, Tech. Rep. NDN-0028, no. November, pp. 1–8, 2016.

Novel Approach for Spatiotemporal Weather Data Analysis

Radhika T V[1], Dr.K C Gouda[2], Dr. S Sathish Kumar[3]
Dept. of Computer Science & Engineering, RNS Institute of Technology, Bengaluru, India[1]
Council of Scientific & Industrial Research (CSIR), Fourth Paradigm Institute (4PI), Bengaluru, India[2]
Dept. of Information Science & Engineering, RNS Institute of Technology, Bengaluru, India[3]

***Abstract*—Massive volumes of multidimensional array-based spatiotemporal data are generated by climate observations and model simulations. The growth in climate data leads to new opportunities for climate studies at multiple spatial and temporal scales. Managing, analyzing and processing of big climate data is considered to be challenging because of huge data volume. In this work multidimensional climate data such as precipitation and temperature are processed and analyzed in the Spark MapReduce Platform, since Spark platform is computationally more efficient than Hadoop-MapReduce Platform of same configuration. In temporal scale monthly and seasonal analysis of climate data has been carried out to understand the regional climate. The prediction of Rainfall on monthly and seasonal time scales is very much important for planning and devising agricultural strategies and disaster management, etc. As the prediction of climate state is very challenging, in this study an attempt is being made for the prediction of the rainfall using the time series analysis in the same framework. As a test case the time series approach such as Auto Regression Integrated Moving Average (ARIMA) has been used for the prediction of rainfall. The proposed approach is evaluated and found to be accurate in the analysis and prediction of climate data and this will surely guide for the researcher for better understanding of the climate and its application to multiple sectors.**

Keywords—Spatiotemporal; big climate data; spark; ARIMA

I. Introduction

Big climate data are preferably provided to scientists for on-demand processing and for analyzing critical problems which may help them to relieve from time consuming computational tasks. Since processing of big climate data requires efficient data management approaches, scalable computing resources and complex parallel computing algorithms, so dealing with this problem is considered to be more challenging task. To address these kinds of challenges, high performance computing technologies have been applied to climate data analysis, modeling and prediction [3]. Many exhaustive big data analytics applications have evolved based on big climate data, also the emergence of various technologies such as Internet of Things (IoT), cloud computing and many advanced Big Data analytics tools have begun to investigate on climate, as well as established various intelligent analytic platforms and new technological advancements have further emphasized its importance and potential impacts on climate science and Big data science development [14][15].

Traditional Big Data techniques are usually incapable of handling large amounts of spatiotemporal data. For example, research has added spatial indexing, spatiotemporal indexing [1] and trajectory analytics features to Hadoop. One of the basic idea of using spatiotemporal data, with respect to large spatial database systems, is the emergence of moving objects [4]. A moving object is a spatial object that varies in geographical position or dimensions over a time period [6]. For Example, Rainfall in one region differs from others; also a river that switches its path over a geologic time scale may be represented as a moving line. Moving region can also be indicated by a hurricane which switches its dimension and geographic position as it evolves [16]. Thus there is a need for High Performance Computing (HPC) environment to process big spatiotemporal data.

The number of cores in HPC environment is persistently getting increased depending on the application’s requirement. Since these applications generate large volumes of spatiotemporal data, that will ultimately be stored and accessed in parallel [15]. The Scientific applications like weather prediction models use standard high-level libraries and data formats such as Network Common Data Format-4 (NetCDF- 4) and Hierarchical Data Format 5 (HDF 5), which will helps to store and operate on the dataset that is situated inside a parallel file system interface. There are various file formats and software libraries in order to reduce the restrictions imposed by plain binary files. Because of which NetCDF file format has been introduced for systematic reading and writing of various kinds of scientific data, mainly for array data. NetCDF file is composed of various kinds of data, which includes BYTE, CHAR, SHORT, LONG, FLOAT, and DOUBLE. The main intention of NetCDF file is to store rectangular arrays of data such as Interactive Data Language (IDL) arrays [12]. NetCDF files are self-descriptive; that is, every file consists of the basic information required to read.

Big spatiotemporal data have gained huge attention in recent years. Analyzing such massive amount of multidimensional data is one of the most common requirements today and processing of this data is considered to be most challenging task. The ability to assess global concerns such as climate change and natural disasters, as well as their influence on different sectors such as agriculture and disease, requires efficient data processing. This is challenging not only because of the large data volume, but also because of the intrinsic high-dimensional nature of the climate data. The emergence of Apache sparks provides quicker solution for big spatiotemporal data analysis and processing speed has reduced drastically compared to the traditional way of processing multidimensional data with multi core processors.

In the proposed work we have used Spark MapReduce Framework for processing of big spatiotemporal data at multiple spatial and temporal scales. In the proposed work we have also considered time series rainfall data, we have read rainfall data (precipitation) of past 11 years (2010-2020) and identified the Box-Jenkins time series seasonal ARIMA approach for prediction of rainfall for Bengaluru region on monthly scales. Seasonal ARIMA model(2, 0, 2) (0, 0, 0) for rainfall was identified the best model to forecast rainfall for next 5 years with confidence level of 76 percent by analysing last 11 year’s data (2010-2020).

Apache Spark is an integrated platform for cluster computing to facilitate efficient big data management and analytics [13]. It is a non-proprietary, distributed computing scheme which enhances the MapReduce framework. Spark system is made of various main modules including Spark core and various high level libraries such as Spark’s MLlib for machine learning, GraphX for graph analysis, Spark Streaming for stream processing and Spark SQL for structured data processing [17]. It functions as a consolidated tool for Machine learning, SQL, Streaming and Graph processing and it supports batch, interactive and stream processing.

Spark is considered to be one of the excellent platform for Data Scientists as it has number of data-centric tools which may assist the data scientists to move forward ahead of the problems that is pertinent in a single machine and also it assist data engineers since it has an integrated method that takes out the need to utilize various special-purpose tools for streaming, machine learning, and graph analytics [13]. More importantly Spark is very essential for researchers, as the platform fosters new opportunities and ideas to design and develop distributed algorithms and also to test their performance in various clusters [9].

The rest of this paper is organized as follows: Section 2 provides overview of various research on Hadoop based approaches to process array-based multidimensional spatiotemporal data; Section 3 presents our proposed Spark-based approach to process multidimensional spatiotemporal data and provides highlights on prediction of rainfall using ARIMA model; Section 4 describes evaluation results of our proposed work by performing sequence of experiments; finally, Section 6 gives summary of the proposed research and envisage on future enhancement.

II. Background Study

Big Data analytics has evolved with advanced opportunities for research, development, business and innovation. It has been identified by four Vs: volume, velocity, veracity and variety and may deliver value via processing of Big Data [2]. The conversion of these four Vs into the 5th (value) is one of the magnificent challenges for processing capacity. The emergence of Cloud Computing as a new standard is to provide computing as a utility service is to deal with various processing needs such as on demand services, pooled resources, elasticity, broad band access and measured services. The capability of delivering computing capacity promotes a possible solution for the conversion of Big Data's 4 Vs into the 5th (value). The continuously increasing volume of big data has accelerated technological developments and practical applications.

Earth is composed of complex dynamic system; as big data analytics works with vast amounts of climate data, it poses greater challenges in climate research than in any other field [3]. Climate change is the present concern throughout globe and also a data-intensive subject, making it one of the main research area for big data experts in recent decades [4][5]. The anomalous growth of climate data makes climate data to be a candidate in the Big Data research. The climate scientists have been exploring on historic data to understand the physics and dynamics, merge millions & billions of daily global observational records and undertake simulations of various climate-change scenarios, all of which leads to huge volumes of data [8].

Extremities in climate such as floods, droughts, and cold and heat-waves may lead to considerable impact on society, ecology and also on the economy globally [7]. Thus spatiotemporal data acquisition, analysis, management and processing are considered to be more important, which will be helpful for various sectoral applications. Spatiotemporal data refers to the data which is connected to both space and time and is considered to be at least 2-dimensional and often 3-dimensional, such that the volume of data gets increased at tremendous speed [8]. Since the general database cannot manage such large volumes of data, there is a need of large database software to play a significant role in the management of spatiotemporal data. Big data is collected from a range of sources, archived, and processed in a variety of computing modes, including cloud computing, mobile computing, edge computing, and wearable computing.

Spatiotemporal data mining is the process of identifying interesting patterns and critical information from spatiotemporal data. Discovering weather patterns, anticipating earthquakes and storms, exposing the progressive history of towns and regions, and identifying global warming trends are the examples of such processes. The unusual rise in spatiotemporal data, combined with the introduction of new technologies, has increased the demand for automatic spatiotemporal knowledge realization. Spatiotemporal data mining techniques are very much essential for many organizations which take decisions based on huge spatiotemporal datasets. As these data are multidimensional in nature, the complexities of such data and their interrelationships create computational and statistical challenges [11].

Researchers of climate science have been exposed to ample of recognized resources of Big climate data for analysis and prediction ,for instance, the NASA Global Climate Change (climate.nasa.gov), the Climate Observing System (GCOS), NASA Center for Climate Simulation (nccs.nasa.gov), Global Earth System Grid Federation (esgf.llnl.gov), the National Center for Atmospheric Research (ncar.ucar.edu), United Nations Global Pulse (unglobalpulse.org), the Climate Data Guide (climatedataguide.ucar.edu), and many other international and national climate analysis and monitoring centers over the world.

Multi-dimensional, array-based data model are mainly used to represent Climate data .The GRIB, HDF and NetCDF are the three most commonly used data formats to store climate data. HDF5/NetCDF4 was mainly developed to enable support for nested structures, ragged arrays, unsigned data types, chunking data structure, and caching techniques which ultimately helps to systematically organize climate science data and to have control over the changing computer models [10]. Meanwhile in order to flexibly use data as multi-dimensional arrays, many software and libraries such as Panoply, h5py, and NetCDF-Java were introduced.

These real time standard software and data formats have added major benefits to store, acquire, examine and exchange climate data. Also there are number of tools available for performing climate data analytics and visualization, one of such tool is Apache Open Climate Workbench, a Python-based tool to carry out interpretations on climate science employing remote sensing rainfall data taken away from various sources and also using climate model outputs.

Since the above mentioned tools and libraries deal with only discrete machines and have restrictions on cloud computing systems, compatibility with HPC and scalability. The absence of proper libraries leads to difficulty in dealing with variety, veracity, format and resolution of Big Climate Data that give rise to a challenge in the emergence of advanced computing technologies.

1) Big climate data management and analytics: In [19] authors have presented a case study supervised by Deutscher Wetterdienst (DWD) which includes storage of array based multidimensional raster data with hands-on exposure on extraction and processing of gridded meteorological data sets. As the big data acquire various challenges such as repositioning, managing and processing with high computational requirements [18]. One of the key resolutions to this is achieved through the database system having the capability of parallel processing and distributed storage. In [19] authors have conducted a study on processing of the multi-temporal satellite image data using SciDB, which is an array-based database mainly used to accumulate, manage and perform computations on such data. The main goal of the proposed work is to provide elastic solution using SciDB to accumulate and execute time series analysis on multi-temporal satellite imagery.

In [21] authors have illustrated the working of SpatialHadoop, It is regarded as one of the first capable open-source MapReduce frameworks to support spatiotemporal data. The working of ST-Hadoop have been illustrated in [20],which has given a support for spatio-temporal data and considered to be one of the first proficient open-source MapReduce framework. In [22] authors have introduced SciHadoopa, a Hadoop plugin that aids scientists in identifying logical queries in data models based on arrays. SciHadoop was used to run queries as map/reduce programs over the logical data model. Authors have shown implementation of a SciHadoop paradigm for NetCDF data and evaluate the performance of five separate optimizations that address the following goals representing an integrated aggregate function query.

2) Time series analysis for rainfall prediction: Time series analysis is a statistical technique that deals with time series data, or trend analysis. Time series data means that data is in a series of particular time periods or intervals. The data is considered to be in three types, such time series data which includes set of observations on the values that a variable takes at different times, Cross-sectional data which is the data made of one or more variables, collected at the same point in time, Pooled data which is a combination of time series data and cross-sectional data. Various research groups attempted to predict rainfall on a seasonal time scales using different techniques. Below we have discussed existing work done related to rainfall prediction using ARIMA.

Climate and rainfall are highly non-linear and complicated phenomena, which require classical, modern and detailed models to obtain accurate prediction. Authors in [23] have considered various statistical models for prediction of rainfall time series data for designing a model, models such as the statistical method based on autoregressive integrated moving average (ARIMA), the emerging fuzzy time series(FST) model and the non-parametric method(Theil’s regression) were used. To evaluate the prediction efficiency, they have used 31 years of annual rainfall data from year 1982 to 2012 of Ibadan South West, Nigeria. ARIMA (1, 2, 1) was used to derive the weights and the regression coefficients, while the theil’s regression was used to fit a linear model. The performance of the model was evaluated using Mean Squared Forecast Error (MAE), Root Mean Square Forecast Error (RMSE) and Coefficient of determination.

To forecast future climatic data, the ARIMA model was utilized. The authors in [24] have proposed ARIMA based daily weather forecasting tool which they have considered as case study for predicting weather of Varanasi. The authors have implemented the ARIMA algorithm in R to create an ARIMA-based weather forecasting tool. The Indian Meteorological Department provided 65 years of daily meteorological data (1951-2015) for this study. The accuracy of the model was calculated according to the root mean square error (RMSE) estimated for each forecasting. They approximated future values for the following fifteen years using ARIMA (2, 0, 2) for rainfall data and ARIMA (2, 1, 3) for temperature data. The root means square error values for rainfall and temperature data were 0.0948 and 0.085, respectively, indicating that the technique functioned correctly. The outcome of this can be further used for the management of solar cell station, agriculture, natural resources and tourism. The error is regarded to be minimal by observing at the values of RMSE, indicating that the ARIMA model has forecasted the data properly.

III. Research Methodology

In the proposed work we have considered Spark MapReduce framework which is considered to be one of the excellent platform for Data Scientists as it has number of data-centric tools which may assist the data scientists to move forward ahead of the problems that is pertinent in a single machine.

1) Data analysis and processing using spark map reduce: As weather data is considered to be multidimensional array based, so in the proposed work, we have considered precipitation and temperature data of Bengaluru region for rainfall prediction and also seasonal weather analysis has been carried out on other states of India. Various experimentation are carried out by reading, analyzing and processing the data. NetCDF data are procured from National Center for Environmental Prediction (NCEP) & India Meteorological Department (IMD) has been used. Following are the work carried out.

- Initially Raw station-level NetCDF based temperature and precipitation data of Bangalore district which is located between 12^o latitude and 77^o longitude has been read in the Google Colab Environment. The data considered for analysis is from Jan 2010-Dec 2020 (11 years data). Data from each year are displayed as a single plot and also 11 years data is also plotted as single graph to analyze past 11 years data and use it for processing to assist in future prediction.
- Mean value has been computed for every year (Jan-Dec) using past 11 years data and plotted as a single point in a graph for analysis.
- Mean value has been computed for all 11 years using Spark MapReduce Platform and plotted as a single graph. This step is considered to be more important as data is effectively processed using spark MapReduce platform for analysis and future weather prediction. The detailed diagram illustrating how the data is processed using spark platform can be seen in (Fig. 1). Following are the steps
  - First step is to import and execute main library files for setting up Spark MapReduce functions in google colab environment.
  - Raw station level precipitation data (pr_wtr) of banglore (from Jan 2010-Dec 2020) is read individually and data frame for each year are created.
  - New data frame is created by adding years as columns (total No. is 12) ie from 2010-2020 and values of corresponding year are placed in the appropriate place and convert the datframe to .csv file.
  - Split Data: As spark MapReduce works by splitting the data and assigning key-value pairs (key is day and value is pr_wtr). In this step, we split the data row wise, and perform read operation using spark.read.option() function. Each row refers to daily data of every year i.e. row 1 is Day 1 data from 2010 to 2020, similarly next row is day 2 data from 2010 to 2020. Same applies till last row which is day 365 from the year 2010-2020. Temporary last column is created in data frame to hold the final row-wise mean value.
  - Map phase: This step computes sum of all the values in each row and calculate 'n' value, where 'n' is No. of columns(2010-2020).we use the formula, n=lit(len(df.column)-1.0) and use the value of it in the next step.
  - Reduce Phase: Row mean is calculated using reduce function of spark ie using reduce ((add,(col(x) for x in df.column[1:]))/n).alias("11 years mean") and the same is displayed.
  - Aggregate Phase: This step Aggregates all mean values, place that in last column created in step 4.
  - Finally display the aggregated precipitation value as a single plot.

2) Seasonal analysis: In the proposed work, we have considered seasonal analysis of temperature and precipitation data of Bangalore to analyze the state of weather during various seasons such as pre-monsoon (March 1-May 31), monsoon (June 1 to September 30), post-monsoon (October 1 to December 31). Various graphs were shown to illustrate season-wise analysis of the weather of particular region such as Bangalore. Comparison of weather status of different cities is undertaken. The results of the same are discussed in Section IV.

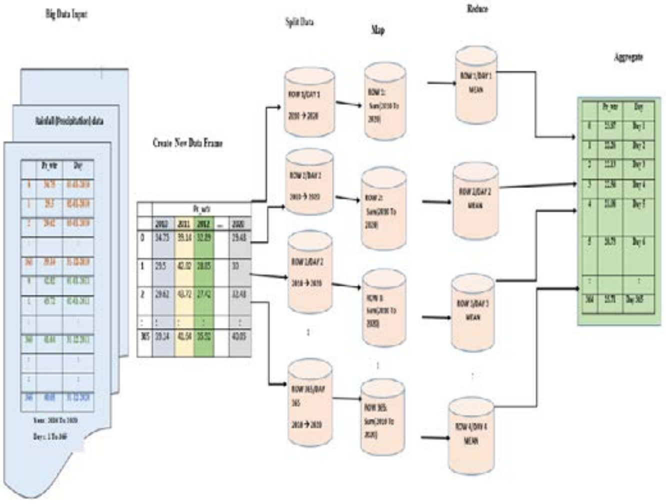


Fig. 1. Spark MapReduce Model to Compute Aggregated Climate Parameters, namely, Precipitation and Temperature.

3) Time series forecasting using ARIMA model: In this study, we have considered past 11 years data and trained them using ARIMA (autoregressive integrated moving average) model. The trained model is used for future forecasting.ARIMA is a class of models that predicts a given time series based on its own past values. An ARIMA model is one where the time series was differenced at least once to make it stationary.

The working principle behind autoregressive (AR) model is that there is a relationship between the present value and the past values. It means that the present value is equal to past values adding with some random value. Moving average (MA) model says that present value is related to the residuals of the past. AR is not capable of forecasting nonlinear data; it can be utilized for data which are linearly related. Using AR and MA together will give best results. But it can be used for stationary weather data and forecasting short term weather. So the proposed work considers ARIMA model which works good for long-term rainfall prediction. We worked on ARIMA (2,0,2) for rainfall data**.**Following steps are used for time series forecasting of rainfall using ARIMA

1) Plot the data.

2) Make the data stationary.

3) Identify the model technique best suited for rainfall forecasting. In the proposed work we have used ARIMA model.

4) Build the model.

5) Compute the mean and Root Mean Squared error (RMSE) value. Use the same for finding accuracy of model.

6) Do the future forecasting based on accuracy of ARIMA model.

Generalized equation used in ARIMA model is as as shown below (1).

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p}\, \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \cdots + \phi_q \varepsilon_{t-q} \qquad (1)$$

Where α is intercept term, β_1 is is the coefficient of lag1 that the model estimates, Y_{t-1} is the coefficient of lag1 that the model estimates.

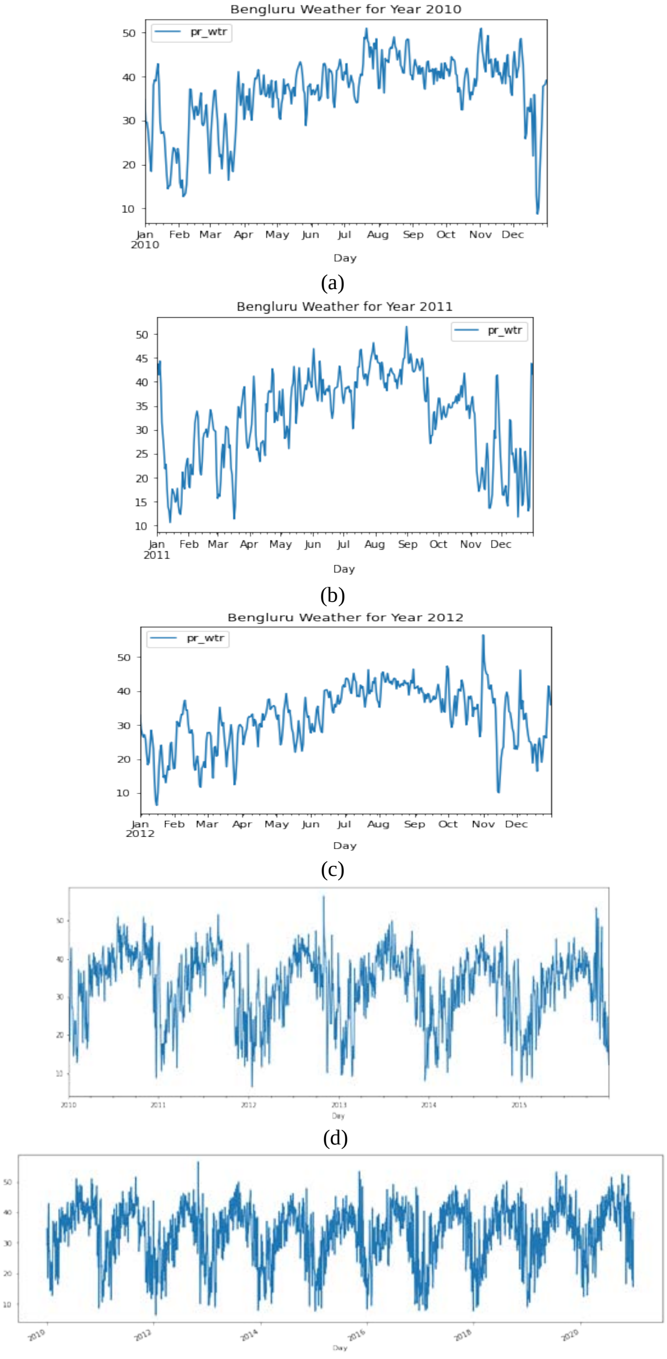
IV. Results and Discussion

The proposed work is executed in Google Colab environment. Python code is used for implementation and necessary libraries were imported. Following are the results. In Fig. 2(a-c) precipitation rainfall data is read individually and plotted as separate graph. Whereas Fig. 2(d-e) shows 5 years and 11 years plot as single graph.

Next we have computed Mean value for every year (Jan-Dec) using 11 years data (2010-2020) and plotted as a single point in a graph for analysis. The same is plotted in line and bar graph as shown in the Fig. 3(a-b).

Mean value has been computed for 11 years (Jan 2010-Dec 2020) using Spark MapReduce Platform and plotted as a single graph. Fig. 4 shows how aggregated mean value is placed in new data frame and shows the final plot after applying to Spark MapReduce.

The Table I shows overview of daily dataset of perceptible water (in mm) for rainfall prediction from the years 2010 to 2020. These daily data of past 11 years has been processed using Spark MapReduce Platform which gives aggregated result as shown in Table II. The same result is used in the analysis and prediction of future rainfall.



(a)

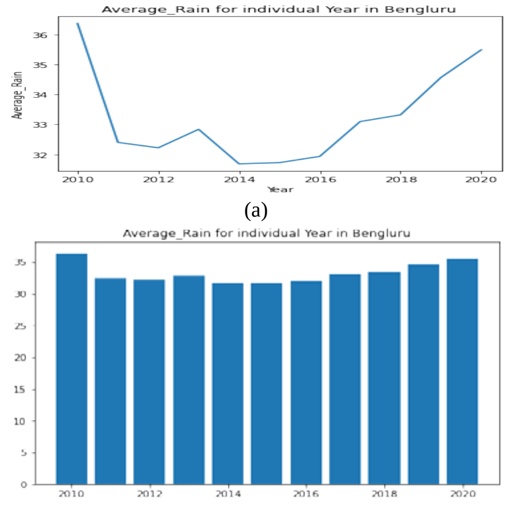
(b)

(c)

(d)

(e)

Fig. 2. (a-e): Raw station-level 11 Years Daily Data (of Bangalore Region) has been Read and Plotted Individually and also as a Single Graph for Comparison.



(a)

(b)

Fig. 3. (a-b): Mean Value for Every Year (Jan-Dec) using 11 Years Data (2010-2020) and Plotted as a Single Point in a Graph for Analysis.

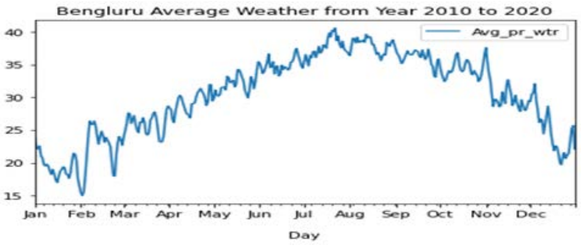


Fig. 4. Final Plot from Aggregated Mean Precipitation Values of 11 Years (2010-2020) using Spark MapReduce Model.

TABLE I. DAILY RAINFALL (PRECIPITABLE WATER) DATASET FROM THE YEAR 2010-2020

Days	Pr_wtr				
	2010	2011	2012	……..	2020
0	34.75	39.14	32.89	……..	29.48
1	29.5	42.82	28.85	……..	30
2	29.62	43.72	27.42	……..	32.48
……	……	……	……	……..	……
364	39.14	41.64	35.92	……..	40.05

TABLE II. AGGREGATED DAILY RAINFALL DATA (PRECIPITABLE WATER) RESULT AFTER PROCESSING 11 YEARS DATASET IN SPARK MAPREDUCE PLATFORM

	Day	Pr_wtr
0	Day 1	23.87
1	Day 2	22.26
2	Day 3	22.13
3	Day 4	22.56
4	Day 5	21.08
5	Day 6	20.73
:	:	:
364	Day 365	25.71

The seasonal analysis of Bangalore weather using temperature data of 2012 has been shown below in Fig. 5. The bar plots shows pre-monsoon, monsoon and post monsoon temperature.

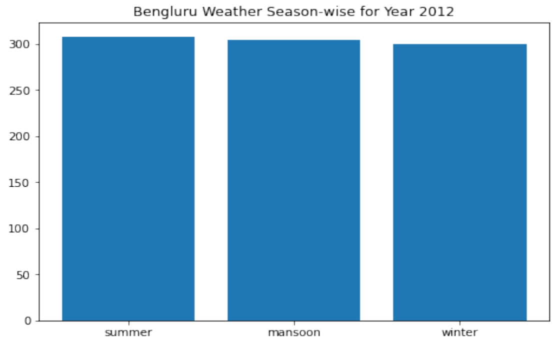
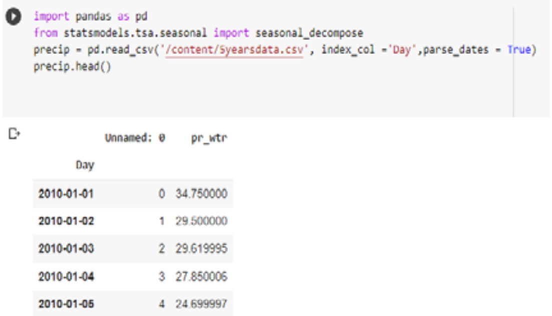


Fig. 5. Seasonal Temperature Analysis of Bangalore for the Year 2012.

The sample first five rows of precipitation dataset (pr_wtr) for the year 2010 is shown in Fig. 6. Time series forecasting is done using ARIMA model. We have worked on ARIMA (2, 0, 2) for rainfall data. Past 11 years rainfall data is trained using the model and the same is used for future prediction. Fig. 7(a-b) shows ARIMA forecasting results. Metrics used for evaluation are Mean error (ME) and Root Mean Squared Error (RMSE). Average error is calculated as shown below in equation (2).

$$\text{Average error} = \text{ME}/\text{RMSE} * 100 \qquad (2)$$

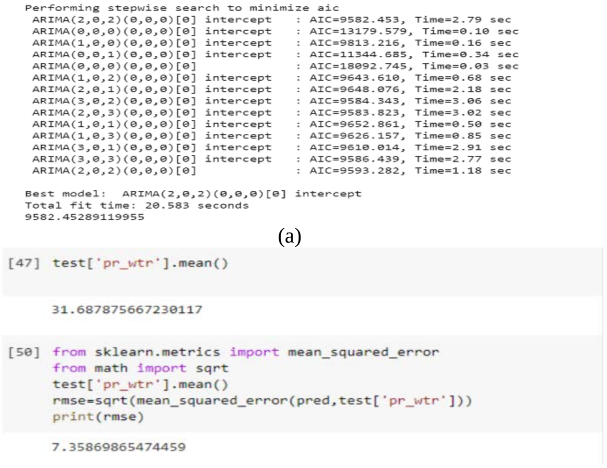
The accuracy for rainfall data considered in our work using ARIMA (2, 0, 2) model is found to be 76.8.



```
import pandas as pd
from statsmodels.tsa.seasonal import seasonal_decompose
precip = pd.read_csv('/content/5yearsdata.csv', index_col ='Day',parse_dates = True)
precip.head()
```

```
            Unnamed: 0     pr_wtr
Day
2010-01-01           0  34.750000
2010-01-02           1  29.500000
2010-01-03           2  29.619995
2010-01-04           3  27.850006
2010-01-05           4  24.699997
```

Fig. 6. Sample View of First Five Rows of Precipitation Dataset for the Year 2010.



```
Performing stepwise search to minimize aic
 ARIMA(2,0,2)(0,0,0)[0] intercept   : AIC=9582.453, Time=2.79 sec
 ARIMA(0,0,0)(0,0,0)[0] intercept   : AIC=13179.579, Time=0.10 sec
 ARIMA(1,0,0)(0,0,0)[0] intercept   : AIC=9813.216, Time=0.16 sec
 ARIMA(0,0,1)(0,0,0)[0] intercept   : AIC=11344.685, Time=0.34 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=18092.745, Time=0.03 sec
 ARIMA(1,0,2)(0,0,0)[0] intercept   : AIC=9643.610, Time=0.68 sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=9648.076, Time=2.18 sec
 ARIMA(3,0,2)(0,0,0)[0] intercept   : AIC=9584.343, Time=3.06 sec
 ARIMA(2,0,3)(0,0,0)[0] intercept   : AIC=9583.823, Time=3.02 sec
 ARIMA(1,0,1)(0,0,0)[0] intercept   : AIC=9652.861, Time=0.50 sec
 ARIMA(1,0,3)(0,0,0)[0] intercept   : AIC=9626.157, Time=0.85 sec
 ARIMA(3,0,1)(0,0,0)[0] intercept   : AIC=9610.014, Time=2.91 sec
 ARIMA(3,0,3)(0,0,0)[0] intercept   : AIC=9586.439, Time=2.77 sec
 ARIMA(2,0,2)(0,0,0)[0]             : AIC=9593.282, Time=1.18 sec

Best model:  ARIMA(2,0,2)(0,0,0)[0] intercept
Total fit time: 20.583 seconds
9582.45289119955
```

(a)

```
[47] test['pr_wtr'].mean()

     31.687875667230117

[50] from sklearn.metrics import mean_squared_error
     from math import sqrt
     test['pr_wtr'].mean()
     rmse=sqrt(mean_squared_error(pred,test['pr_wtr']))
     print(rmse)

     7.35869865474459
```

(b)

Fig. 7. (a-b): Preview and Results of ARIMA (2 ,0, 2) Model.

V. Conclusion

Vast amounts of climate data are being generated rapidly by satellite observations and numerical climate models. Agriculture, tourism, water, electricity, wildfire management, and other sectors are all require climate data. The utility of climatic data depends on timely analysis. Existing technologies, such as Apache Hadoop, which are based on the idea of breaking problems down into smaller chunks and solving them on a cluster of commodity servers, have emerged as a possible solution for analysing huge climate datasets. Apache Spark has recently emerged as a viable alternative to Hadoop's disk-based architecture. The proposed work considers analysis and processing of big spatiotemporal data using Spark MapReduce platform. Multidimensional NetCDF based precipitation and temperature data from NCEP and CSIR-4PI are considered for analysis. Analysis shows that Spark platform is computationally more efficient (double the No. of times) than Hadoop - MapReduce Platform of same configuration. Monthly and seasonal analysis of climate data has been carried out. Time Series prediction approach such as ARIMA (2,0,2) model was used for forecasting future rainfall of Bangalore region, results shows that ARIMA performs well for long term weather prediction. Performance analysis of the model has been carried out using NetCDF data of NCEP and CSIR-4PI Bangalore.

References

[1] Z. Li, F. Hu, J. L. Schnase, D. Q. Duffy, T. Lee, M. K. Bowen, and C. Yang. A spatiotemporal indexing approach for efficient processing of big array-based climate data with MapReduce, International Journal of Geographical Information Science, pages 17–35, 2017.

[2] Chaowei Yang, Manzhu Yu, Fei Hu,Yongyao Jiang,Yun Li, Utilizing cloud computing to address big geospatial data challenges, Journal of Computers, Environment and Urban Systems, 2016, http://dx.doi.org/10.1016/j.compenvurbsys.2016.10.010.

[3] James H. Faghmous and Vipin Kumar, A big data guide to understanding climate change: The case for theory-guided data science, Journal of Big Data, Vol. 2, No. 3, Sep 2014, Pages 155–163, https://doi.org/10.1089/big.2014.0026.

[4] Markus Götz, Christian Bodenstein, Matthias Richerzhagen, Gabriele Cavallaro. On Scalable Data Mining Techniques for Earth Science, Procedia Computer Science, December 2015, Volume 51, Pages 2188–2197.

[5] Jinsong Wu, Song Guo, Jie Li,Deze zeng, Big data meet green challenges: Greening big data. IEEE Systems Journal, Volume: 10, Issue: 3, 19 May 2016, Pages 873 – 887.

[6] Ralf Hartmut Guting, M. H. Bohlen, Martin Erwig, Christian S. Jensen, Nikos A. Lorentzos, Markus Schneider, and Michalis Vazirgiannis, A foundation for representing and querying moving objects, ACM Transactions on Database Systems (TODS), Vol. 25, No. 1, March 2000, Pages 1–42.

[7] Sebestyén Viktor, Czvetkó Tímea, Abonyi János, The Applicability of Big Data in Climate Change Research: The Importance of System of Systems Thinking, Frontiers in Environmental Science, Volume 9, March 2021,DOI:10.3389/fenvs.2021.619092.

[8] Yang C., Clarke K., Shekhar S., Tao C.V, Big Spatiotemporal Data Analytics: a research and innovation frontier, International Journal of Geographical Information Science, April 2020, https://doi.org/10.1080/13658816.2019.1698743.

[9] Fei Hu, Chaowei Yang, Daniel Q. Duffy, Michael Bowen, Weiwei Song, Tsengdar Lee, Mengchao Xu and John L. Schnase, ClimateSpark: An in-memory distributed computing framework for big climate data analytics, Journal of Computers and Geosciences, March 2018, Pages 154-166, https://doi.org/10.1016/j.cageo.2018.03.011.

[10] Christopher Bartz, Konstantinos Chasapis, Michael Kuhn, Petra Nerge & Thomas Ludwig, A Best Practice Analysis of HDF 5 and NetCDF- 4 Using Lustre, International Conference on High Performance Computing, ISC High Performance 2015: High Performance Computing, volume 9137, Pages 274-281.

[11] Gowtham Atluri,Anuj Karpatne, Vipin kumar, Spatio-Temporal Data Mining: A Survey of Problems and Methods, ACM Computing Surveys, Volume 51, Issue 4, Article No.: 83, July 2019 Pages 1–41,https://doi.org/10.1145/3161602.

[12] R. Rew, G. Davis, NetCDF: an interface for scientific data access, Volume: 10, Issue: 4, July 1990, pages: 76 – 82, DOI: 10.1109/38.56302.

[13] Salman Salloum, Ruslan Dautov, Xiaojun Chen, Patrick Xiaogang Peng, Joshua Zhexue Huang, Big data analytics on Apache Spark, International Journal of Data Science and Analytics, Springer International Publishing Switzerland 2016, Pages 145-164.

[14] Abdul Salam, Internet of Things for Sustainable Human Health, Book chapter in Internet of Things for Sustainable Community Development. Internet of Things, Springer, January 2020, Pages 217-242, https://doi.org/10.1007/978-3-030-35291-2_7.

[15] Pankaj Mudholkar and Megha Mudholkar, Internet of Things (IoT) and Big Data: A Review, International Journal of Management, Technology and Engineering, Volume 8, Issue XII, December 2018, ISSN NO: 2249-7455, Pages 5001-5007.

[16] Mark McKenney, Niharika Nyalakonda, Jarrod McEvers, Mitchell Shipton, Pyspatiotemporalgeom: A Python Library for Spatiotemporal Types and Operations, Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, October 2016, Article No.: 93, Pages 1–4.

[17] R. Rew, G. Davis, NetCDF: an interface for scientific data access, IEEE Journal of Computer Graphics and Applications, Volume: 10, Issue: 4, July 1990, Pages 76 – 82.

[18] Dimitar Misev, Peter Baumann, Jürgen Seib, Towards Large-Scale Meteorological Data Services: A Case Study, Journal of Datenbank Spektrum- Springer, Volume 21,Issue 1,Pages183–192, 22nd September 2012.

[19] A. Joshi, E. Pebesma, R. Henriques, M. Appel, SCIDB Based Framework For Storage And Analysis Of Remote Sensing Big Data, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-5/W3, Capacity Building and Education Outreach in Advance Geospatial Technologies and Land Management,pp.10–11 December 2019, Dhulikhel, Nepal.

[20] Louai Alarabi, Mohamed F. Mokbel, A Demonstration of STHadoop: A MapReduce Framework for Big Spatiotemporal Data, proceedings of the VLDB Endowment, Vol. 10, No. 12, August 2017.

[21] Ahmed Eldawy, Mohamed F. Mokbel, Demonstration of SpatialHadoop: An Efficient MapReduce Framework for Spatial Data, proceedings of the VLDB Endowment, Volume 6, Issue 12, August 2013, Pages 1230-1233, https://doi.org/10.14778/2536274.2536283.

[22] Joe B. Buck, Noah Watkins, Jeff LeFevre, Kleoni Ioannidou, SciHadoop: Array-based query processing in Hadoop, Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, November 2011, Article No.: 66, Pages 1–11 https://doi.org/10.1145/2063384.2063473.

[23] Timothy Olatayo, A. I. Taiwo, Statistical Modelling and Prediction of Rainfall Time Series Data, Global Journal of Computer Science and Technology: Interdisciplinary, Volume 14, Issue 1 Version1.0, 2014, Online ISSN: 0975-4172 & Print ISSN: 0975-4350.

[24] Nikita Shivhare, Atul Kumar Rahul, Shyam Bihari Dwivedi and Prabhat Kumar Singh Dikshit, ARIMA based daily weather forecasting tool: A case study for Varanasi, Journal Mausam 70(1), January 2019, Pages 133-140.

Prediction Models to Effectively Detect Malware Patterns in the IoT Systems

Rawabi Nazal Alhamad[1]
Department of Computer Science and Technology
Jouf University, Al-Jouf, Saudi Arabia

Faeiz M. Alserhani[2]
Department of Computer Engineering and Networks
Jouf University, Al-Jouf, Saudi Arabia

***Abstract*—The Widespread use of the Internet of Things (IoT) has influenced many domains including smart cities, cameras, wearables, smart industrial equipment, and other aspects of our daily lives. On the other hand, the IoT environment deals with a massive volume of data that needs to be kept secure from tampering or theft. Detection of security attacks against IoT context requires intelligent techniques rather than relying on signature matching. Machine learning (ML) and Deep Learning (DL) approaches are efficient to detect these attacks and predicting intrusion behavior based on unknown patterns. This study proposes the application of five deep and ML techniques for identifying malware in network traffic based on the IoT-23 dataset. Random Forest, Catboost, XGBoost, Convolutional Neural Network, and Long Short-Term Memory (LSTM) models are among the classifiers utilized. These algorithms have been selected to provide lightweight security systems to be deployed in the IoT devices rather than a centralized approach. The dataset was preprocessed to remove unnecessary or missing data, and then the most significant features were extracted using a feature engineering technique. The highest overall accuracy achieved was 96% by applying all classifiers except LSTM which recorded a lower accuracy.**

Keywords*—*Internet of Things (IoT); malware deletion; random forest; Catboost; convolutional neural network; long short-term memory (LSTM); XGBoost

I. Introduction

The Internet of Things (IoT) refers to the billions of connected physical devices through the Internet, for global storage and data exchange [1]. Recently, the IoT has been involved in a variety of fields in our daily life, including smart cities, cameras, wearables, smart industrial equipment, household appliances, medical devices, and even nuclear reactors [2]. The infrastructure of the IoT devices has limited storage, hardware, and battery life, making the sophisticated or standard security algorithms difficult to be applied in such a domain with limited resources it has. Furthermore, the IoT environment deals with a large amount of data, making it subject to botnets, firmware hijacking, distributed denial of service attacks, eavesdropping, the man in the middle, and other attacks. The network security of IoT devices is considered more critical compared to network security due to the large number of attacks, its small size, and multiple vulnerabilities of IoT tools [3]. By 2025, it is expected that 41,600 million IoT devices will have been shipped around the world. According to the 2022 SonicWall Cyber Threat Report [4], malware decreased somewhat in 2021, indicating a third consecutive year of decline and a seven-year low. An increase in attacks during the second half of 2021 nearly wiped out the 22 percent decline in malware that researchers observed at the midpoint of the year, reducing the total decrease for 2021 to just 4 percent, where 2022 Global Cyberattack was registered approximately 60,1 million IoT Malware attacks, whose volume increased by 6 percent in 2021. This growth indicates a plateau compared to the previous two years, during which these attacks increased by 218 percent and 66 percent, respectively.

The research question behind this work is “Is it possible to enhance the accuracy of malware and benign detection in the IoT environment based on deep and ML techniques using a standard dataset that holds network traffic information? And what is the potential of deploying these techniques in end-systems?” Hence, to solve this question, five different deep and machine learning approaches are suggested to be applied to the IoT-23 dataset. The IoT-23 dataset has 14 labels for 20 malicious and 3 benign captures and many features; therefore, to reduce the computational cost of using the IoT-23 dataset, only the malware captures were explored, and only ten labels were analyzed.

The main purpose of this study is to analyze traffic traces and network behavior of IoT devices by using the 23-IoT dataset, which consists of attributes from network traffic based on different protocols to identify malware. After that, using preprocessing stage and feature engineering to extract the most significant features and thereby reduce the dataset's dimensionality. Then applies classification techniques that include Random Forest, Catboost, CNN, LSTM, and XGBoost algorithms, to detect malware and benign traffic, which helps in developing a robust intrusion system capable of detecting different types of attacks. Finally, to accomplish the evaluation and better prediction, the classifier models are subjected to cross-validation, optimization, and comparison.

The following is the structure of this research study: Section 2 depicts the related works by applying deep and machine learning techniques to detect malware activities based on the IoT-23 dataset. Section 3 outlines the suggested methodology in this research, while Section 4 discusses the various machine learning and deep learning models implemented. Moreover, Section 5 discusses the results of each model. Finally, the conclusion and future work.

II. Related Work

The evolution of the sophisticated IoT environment has necessitated the development of malware and benign system detection based on the recent deep and ML techniques, as detailed in the previous section. Many researchers have used various deep and ML models to achieve network traffic analysis to address this issue. For example, in [5], the authors implemented deep and ML algorithms, namely, RF, Naive Bayes (NB), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and AdaBoost (ADA). The dataset used in this work was the IoT-23 dataset and the best accuracy of detection was achieved by using the RF model with a value of 99.5%.

In [6], the researchers applied four algorithms to the IoT-32 dataset, including CNN, Decision Trees, Naive Bayes, and SVM. All IoT-23 dataset's labels were analyzed in this work the highest accuracy achieved was 73% by using the decision tree model, CNN achieved 69.35%, and SVM attained 69% of accuracy. Based on the same dataset, another author was analyzing all labels of the dataset by using Decision Trees, RF, Naive Bayes, SVM, AdaBoost, XGBoost, CNN, and Multi-Layer Perceptron models [7]. The highest accuracy achieved was 74% by-using Decision Trees, RF, and MLP models.

On the other hand, the authors in [8] achieved a better detection rate near 100% of the f1-score metric; however, only four malware types (Hide and Seek, Torii, Mirai, and Trojan) were involved and Okiru malware was not analyzed. Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Isolation Forest (iForest), Local Outlier Factor (LOF), and Deep Reinforcement Learning (DRL) classifiers were used to do binary classification and multiclassification.

On another trend, a DL ensemble for network malware detection was proposed based on IoT-23, LITNET-2020, and NetML-2020 datasets [9]. Deep Neural Network (DNN) and Long Short-Term Memory (LSTM), and a meta-classifier (i.e., logistic regression) were applied. The method employed a two-step process for the detection of network anomalies to improve the capabilities of the suggested methodology. For the feature engineering challenge in the first stage, data pre-processing, a Deep Sparse AutoEncoder (DSAE) was used. For classification in the second phase, a stacking ensemble learning strategy was applied. The proposed method was evaluated on IoT-23, LITNET-2020, and NetML-2020 datasets, and only Benign, Mirai, Attack, PartOfAHorizontalPortScan, and C&C malware types are classified in this work. The classification type that was applied by the authors is binary classification to recognize normal from abnormal attacks and the accuracy achieved is around 100%.

In [3], the authors have built and performed IoT anomaly detection systems based on actual IoT-23 large data for identifying attacks based on artificial NN like Convolutional NN (CNN), Recurrent Neural Networks (RNN), and Multilayer Perceptron (MLP). As a consequence, Convolutional Neural Networks outperform Multilayer Perceptron and Recurrent Neural Networks in IoT anomaly detection, with a metric accuracy score of 0.998234 and a minimal loss function of 0.008842. The Mirai, DoS, NScan, Normal, and MITM_ARP attacks are classified in this work.

An anomaly-based intrusion detection model to detect malware in IoT network traffic was created and implemented on IoT-23, IoT Network Intrusion, BoT-IoT, and MQTT-IoT-IDS2020 datasets [10]. A multiclass classification model was created using CNN (Convolutional Neural Network) models and then uses to accomplish binary and multiclass classification via the transfer learning principle. Three model architectures were suggested by the authors, namely, 1D, 2D, and 3D CNN models. The utilized multiclass classifier not only can classify 15 different types of attacks but also efficiently differentiate them from normal network data [11]. The authors included seven attacks and one normal from the IoT-32 dataset: Normal, Attack, C&C, FileDownload, HeartBeat, Okiru, Port Scan, and Torii attacks. The proposed models achieved high values in the used performance metrics, where all models reached more than 99.89% for accuracy, precision, recall, and f1-score metrics.

The authors in [12] used Bidirectional Generative Adversarial Networks (BiGAN) and Adversarial Autoencoders (AAE) to detect malware in network traffic based on the full IoT-23 dataset version. The proposed models outperformed traditional ML such as RF, by getting an f1-score of 99. However, not all labels were analyzed in this work, only nine malware types were classified. Sahu, Amiya Kumar, et al. [13] proposed a DL-based classifier for detecting IoT attacks. The CNN model is used to learn the IoT features, and then an LSTM-based classifier is used to classify them. The model was applied to eight labels on the IoT-23 dataset, which included Command and Control (C&C), Distributed Denial of Server (DDoS), File Download, Heart Beat, Part of A Horizontal Port Scan, Mirai, Torii, and Okiru. A 96% accuracy rate in detecting malicious devices was achieved but the Benign label was ignored. Dartel, Bram. The author in [14] suggested a ML technique comprising Decision Tree, RF, Support Vector, and an ESP32. The sub-labels of the IoT-23 dataset are judged irrelevant because their methodology is focused on malware detection rather than malware classification. The result was running on an IoT device that really works as an IoT device, so it was easy to run the device next to the malware detection algorithm.

Using a similar size dataset, they divided it into two separate datasets. In addition, the data were randomly dispersed throughout them in order to limit the number of multiclass labels, hence the metrics showed a significant disparity in class size. Following this, they used the following ML techniques: RF, Naive Bayes, Support Vector Machine, and Decision Tree. 99.5 percent of the accuracy of the RF algorithm produced the best results [15]. They suggested using ML techniques like Support Vector Machine, Decision Tree, and RF to classify malware attacks like DDoS attacks, and also, a Principal.

Component Analysis (PCA) was used to reduce the number of dimensions. The results of PCA were measured against what would have happened if PCA hadn't been used. When PCA was used, the algorithm ran much faster with fewer features than it did when PCA wasn't used. When it comes to classifying attacks, Decision Tree and RF are better than SVM [16].

TABLE I. COMPARISON OF RELATED WORK

Author & year	Study Name	Method	Accuracy	Features selected	Notes
Ullah, Imtiaz Mahmoud, Qusay H[11] 2021	Design and development of a DL-based model for anomaly detection in IoT networks	1D, 2D, and 3D CNN	99.89%	Normal, Attack, C&C, FileDownload, HeartBeat, Okiru, Port Scan, and Torii attacks	They get high accuracy with three models by CNN but for 7 attacks classes with normal class
Abdalgawad, N Sajun, [21] 2021	Generative DL to detect Cyberattacks for the IoT-23 Dataset	(BiGAN), (AAE), RF	getting an f1-score of 99.	9 attacks	not all labels were analyzed
Sahu, Amiya Kumar Sharma, [13] 2021	IoT attack detection using hybrid DL Model	CNN, LSTM	96%	eight labels on the IoT-23 dataset, which included Command and Control (C&C), Distributed Denial of Server (DDoS), File Download, Heart Beat, Part Of A Horizontal Port Scan, Mirai, Torii, and Okiru	8 attacks and Benign label was ignored
Dr. R. Thamaraiselvi and S. Anith [14] 2021	Malware detection in IoT devices using ML	DT, RF, SVM, and an ESP32	Training model on Iot-23. Testing was running on an IoT device that works as an IoT device	Sub Labels of IoT-23	
R. Thamaraiselvi [15]- 2020	Attack and Anomaly Detection in IoT Networks using ML	RF, Naive Bayes, Support Vector Machine, and Decision Tree	99.5	Split the dataset into two parts	
D. Nanthiya [16] 2021	SVM Based DDoS An IoT Using Iot-23 Botnet Dataset	SVM, DT , RF	SVM is higher	Principal Component Analysis (PCA) was used to reduce the number of dimensions	

Authors in [17] proposed a method for Malware Detection in Fog Layer. This method has three important phases for pre-processing: feature extraction, feature selection to reduce the number of features, and classification. Convert the binary files to hexadecimal using HexDump, segmentation into 4-gram, features selection and reduction using Gain Ratio; decision tree and Component Analysis (PCA). Finally, apply decision tree classifier. The accuracy was Acc. 96.7. The authors [18] have presented a new method based on ML for detecting Mirai malware "NBaIoT" dataset, which data consist of features infected by the Mirai Malware, is used in that study. The Cross-Validation technique has been used for data splitting to overcome overfitting, and the experiment was conducted using ANN. The achieved accuracy is 92.8%. The Opcode dataset has been used in this research; it consists of 70,140 normal and 69,860 malicious malware. The IoT Device dataset's benign or malignant input is classified using a deep neural network (DNN). The obtained accuracy reached 99.7 % [19]. Table I summarize the related works and provide a comparison between different approaches.

III. METHODOLOGY

Generally, the research methodology has been designed for detecting about 10 IoT malware types which will be explained in detail in this section. To detect IoT malware attacks, five ML and DL algorithms are implemented based on the IoT-23 dataset, namely, RF, Catboost, Convolutional Neural Network, Long Short-Term Memory (LSTM), and XGBoost. The proposed methodology has five stages to be accomplished to evaluate the algorithms: data collection, pre-processing, feature selection, training and testing, and classification stages as shown in Fig. 1.

A. Data Collection

To use machine learning techniques, a dataset with a large number of samples that have been contextualized and labeled correctly is necessary. This section gives a quick overview of the chosen dataset in this work. The IoT-23 dataset is chosen since it provides a large dataset with twenty-three captures of various IoT network traffic that include three benign and twenty malware traffic captures [20]. It is a modern dataset that consists of more than one million network traffic of IoT devices and was first published by the Stratosphere Laboratory in January 2020, with captures spanning the years 2018 to 2019. The main goal of creating this dataset is to create a large dataset under real circumstances for researchers that h valid malware and benign traffic labels in order to apply machine learning algorithms to enhance intrusion detection in IoT environments.

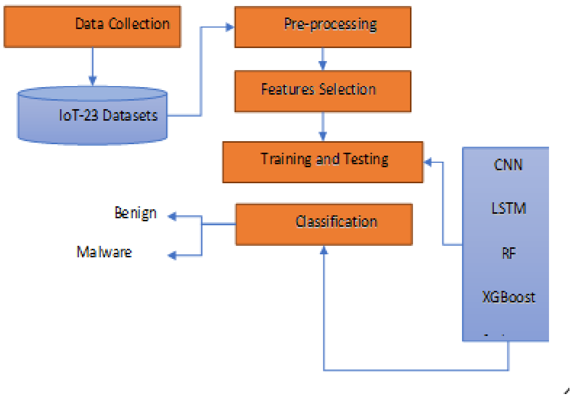


Fig. 1. The Proposed Methodology in this Work.

This dataset has 14 labels for 20 malicious and 3 benign captures that include Part-Of-A-Horizontal-PortScan, DDoS, Attack, FileDownload, Okiru, Benign, C&C-HeartBeat-FileDownload, C&C, C&C-FileDownload, C&C-HeartBeat, C&C-Mirai, C&C-FileDownload, Okiru-Attack, and C&C-Torii. Despite this, the dataset contains 21 feature properties that determine the feature of the connections, one of which is the class label, as shown in Table II. Some of the features are nominal, and some features have time-stamp values, while others are quantitative. In this work, only the malware captures were investigated to reduce the computational cost of using the full version of the IoT-23 dataset, as well as only ten labels were analyzed: Part-Of-A-Horizontal-PortScan (753565 rows), DDoS (138777), C&C-Mirai (1), C&C-HeartBeat (341), C&C (15100), Attack (3915), Benign (195270), FileDownload (13), C&C-FileDownload (43), C&C-Torii (30), and C&C-HeartBeat-FileDownload (8) labels.

TABLE II. THE DESCRIPTION OF THE IMPORTANT FEATURES OF IOT-23 DATASET [9]

Feature No.	Feature Name	Data Type	Feature Description
1	Ts	int	Timestamp of the capture.
2	Uid	str	The capture's Unique ID.
3	id.orig_h	str	Originating IP where the attack happened.
4	id.orig_p	int	Source port used by the responder.
5	id.resp_h	str	The destination IP address of the device on which the capture happened.
6	id.resp_p	int	Destination port used from the response from the device on which the capture happened
7	Proto	str	Transaction or Network protocol
8	Service	str	Application protocols such as DNS, FTP, HTTP, SMTP, SSH, etc.
9	Duration	float	The overall duration of the transmission between device and attacker
10	orig_bytes	int	The transaction bytes from source to destination.
11	resp_bytes	int	The transaction bytes from destination to source.
12	conn_state	str	Represents the current connection state
13	local_orig	bool	The connection is locally initiated.
14	local_resp	bool	The response is locally initiated.
15	missed_bytes	int	The number of missing bytes of a transaction
16	History	str	The connection state's history.
17	orig_pkts	int	The total packets being sent to a device.
18	orig_ip_bytes	int	The total bytes being sent to a device.
19	resp_pkts	int	The total packets being sent from a device.
20	resp_ip_bytes	int	The total bytes being sent from a device.
21	Label	str	Type of capture: Benign or malicious, alongside with Type of the malicious capture

B. Pre-processing Stage

Data pre-processing is the process that transforms raw data into a form that can be read, accessed, and analyzed. The pre-processing stage is of major importance, to ensure or enhance the total performance or accuracy of any system before applying the machine learning algorithms. In this work, the full IoT-23 dataset is used that has around 20GB in size. When we extract this file into the local hard disk, in windows 11, 23 folders are created; however, only 20 folders are utilized that are relevant to malware captures. To read the data from each folder, a 'read_table' function from Pandas is imported to extract all data from 'conn.log.labeled' file, this function is applied in the Jupyter platform that supports python programming language. Therefore, 20 variables are created to read all data from each folder, and then the 'concat' function from Pandas is applied to combine all these variables for creating a Dataframe that can save data to a CSV file using 'to_csv'. But before saving such data to the CSV file, we need to change the labels for some rows of the extracted data. For example, some rows have a '-Malicious PartOfAHorizontalPortSca' label, while others have '(empty) Malicious PartOfAHorizontalPortScan', so all these labels denote one attack and will be changed to the 'PartOfAHorizontalPortScan' label; therefore, we did the same step for all other labels for creating unique labels.

C. Features Selection

The process of selecting the features that have the greatest impact on the prediction outcomes is of major importance to increase the overall accuracy. Therefore, in this section, effective steps will be explained in detail. After the pre-processing stage, all data is stored in a CSV file named 'iot23_combined.csv', and this file is loaded to a Dataframe by using the 'read_csv' Panda's function. After that, the first feature is removed since it represents the number of rows in the IoT-23 datasets. On the other hand, all labels will be analyzed in this work except the 'Okiru' label, because of the computational cost of the large size of the IoT-23 dataset, as well as we found that some researchers have done malware detection by using only a few labels not all of them [8], [9].

Furthermore, in order to delete data that is not related to the 'label' column in this dataset, a correlation matrix is applied to all features, as shown in Fig. 2. The correlation matrices show the strength of correlation by using a scale from dark blue to yellow. The negative correlation is represented by dark-blue, while yellow represents the positive correlation, and green denotes to the weak correlation. From the figure, the yellow color indicates that the features are strongly correlated while the dark blue indicates that the features are weakly correlated. Furthermore, the repeated features should be removed when any exist by applying the 'get_duplicate_features' function from the 'fast_ml' library, and now the data is ready for the training and testing stage.

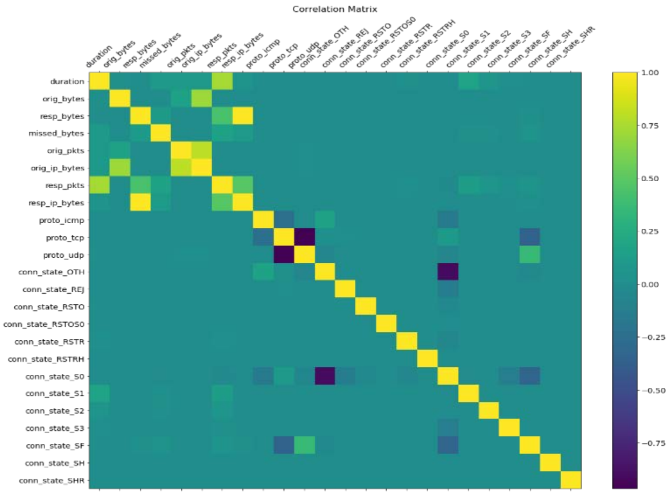


Fig. 2. The Correlation Matrix for All IoT-23 Dataset's Features.

D. Training and Testing

The process of training and testing the proposed machine learning models will be explained in this stage. This process depends on the prepared dataset in the previous stage to train and test five machine learning algorithms (CNN, LSTM, RF, XGBoost, and Catboost) until these models can distinguish the malware from the benign captures. However, there are a few steps that should be taken into consideration such as splitting the data into test and train data by applying the 'train_test_split' function from the Sklearn library, which is a function to split the matrices into random train and test subsets. Sklearn is a free python library, and being created to perform some techniques in the machine learning field such as classification, regression, and grouping. Applying the 'train_test_split' function is very important for an unbiased evaluation of all suggested machine learning models and checking the final accuracy not only in the training data but also in the test data that have not been seen or trained before in such models. In this work, the size of training data is 80%, while the test data is 20% of all the prepared data. After all these steps, all the models are ready to be trained and tested with different parameters depending on the requirements of the model being run at the time of execution.

E. Classification

Classification is a supervised learning perception that is responsible forsplittings data into separate classes, and can be applied in various classification issues such as facial detection, speech recognition, handwriting recognition, and document categorization. In this work, after the proposed models are trained on the IoT-23 dataset, and then a set of data isolated from these models are used to check the accuracy of the trained models to correctly separate all data according to their labels. Accuracy, Precision, F1-Score, and Recall are the measures used to evaluate the algorithms' efficiency, and their definitions are as follows:

- Accuracy: is the percentage of correctly labeled classes in relation to the total number of classes and is given by the next equation:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Precision: how many of the positive class classifications made by the model are correct? and is given by the next equation:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

- Recall or Sensitivity: how many of the positive class scenarios with expected values are correct? and is given by the following equation:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

- F1-score: is a harmonic average that combines precision and sensitivity into one measure, and the following equation shows how to calculate this measure:

$$F1-score = \frac{2*Precision*Recall}{Precision+Recall} \tag{4}$$

Where TP (True-Positive) means the model classifies malware cases as positive (malware cases classified as malware correctly). Whereas FN (False-Negative) represents mistakenly classified malware as negative (malware classified as normal or benign but the truth is that malware has existed). Moreover, TN (True-Negative) denotes correctly current cases classified as negative or no malware that has existed in the current cases. Finally, FP (False-Positive) means mistakenly classified malware as positive (malware classified as malware but the truth is that no malware exists).

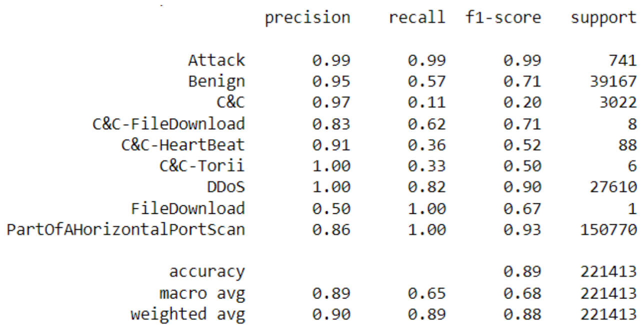
IV. Experimental Results

In this section, various experiments were carried out intending to check the accuracy of detection of the proposed models in this work. To effectively identify the IoT malware and benign captures, Random Forest (RF), CatBoost, XGBoost, LSTM, and CNN were implemented. The measure performances used in this work include confusion matrix, classification report (precision, recall, f1-score, accuracy), and ROC curves, for comparing the different performance of the proposed models.

A. Random Forest Classifier's Experiments

As mentioned in the implementation section, a random forest classifier is used because of its accuracy is higher, due to this classifier takes the final prediction based on the forecast from each tree, not from a single decision tree. The random forest achieved the highest detection accuracy for differentiating malware and benign captures with a value of 89% as portrayed in Fig. 3. Besides, in this figure, three measure metrics ((precision, recall, f1-score) are appeared for each label, while the overall accuracy is presented to show the total detection accuracy for all labels included in this work. The highest precision attained is for C&C-Torii and DDoS with a value of 100% for both of them, the highest recall was 100% for FileDownload and PartOfAHorizontalPortScan, and the highest f1-score was 99% and 93% for Attack and PartOfAHorizontalPortScan, respectively. The f1-score is more important than precision and recall because it can assist balancing the metric between positive and negative samples. Additionally, two malwares (C&C-HeartBeat-FileDownload, C&C-Mirai) have not appeared in the classification report because the small number of samples they have: 1 for C&C-

Mirai and 8 samples for C&C-HeartBeat-FileDownload. Furthermore, this the figure has micro-averaging and macro-averaging curves, where the micro-averaging is used to score each prediction equally and the macro- averaging is used to examine the overall performance of the classifier based on the most common class labels.



```
                          precision    recall  f1-score   support

                  Attack       0.99      0.99      0.99       741
                  Benign       0.95      0.57      0.71     39167
                     C&C       0.97      0.11      0.20      3022
         C&C-FileDownload      0.83      0.62      0.71         8
            C&C-HeartBeat      0.91      0.36      0.52        88
               C&C-Torii       1.00      0.33      0.50         6
                    DDoS       1.00      0.82      0.90     27610
             FileDownload      0.50      1.00      0.67         1
PartOfAHorizontalPortScan      0.86      1.00      0.93    150770

                accuracy                           0.89    221413
               macro avg       0.89      0.65      0.68    221413
            weighted avg       0.90      0.89      0.88    221413
```

Fig. 3. The Classification Report of Random Forest Model.

To draw a ROC (receiver operating characteristic curve) curve for these labels, we need to encoded them to integers ranging from 0 to 8: Attack=0, Benign=1, C&C=2, C&C-FileDownload=3, C&C-HeartBeat=4, C&C-Torii=5, DDoS=6, FileDownload=7, and labels PartOfAHorizontalPortScan=8. These labels were numbers according to their appearance in the classification report in Fig. 3. The ROC curve shows the performance of the classification of the model, as shown in Fig. 4. Additionally, as we see this graph has classes that range from 0 to 8, where class 0 = Attack, class 1= Benign, and so on. The best ROC curves area values were 100% for the Attack and FileDownload.

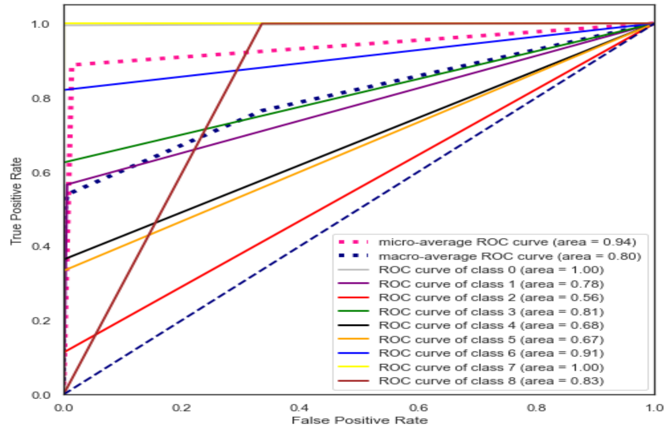
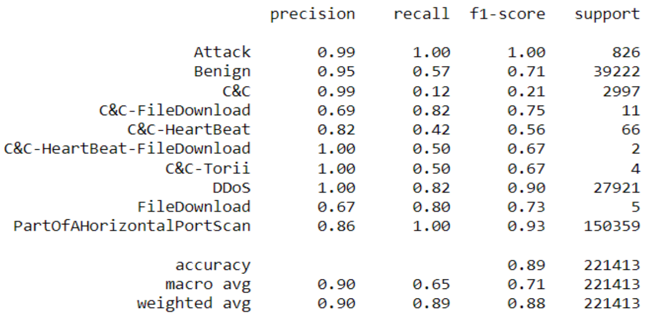


Fig. 4. The ROC Curve of Random Forest Model.

B. *XGBoost Classifier's Experiments*

The XGBoost classifier achieved the same total accuracy as the random forest with a value of 89% as shown in figure 5. The highest precision attained is for C&C-Torii, C&C-HeartBeat-FileDownload, and DDoS with a value of 100% for all of them, the highest recall was 100% for Attack and PartOfAHorizontalPortScan, and the highest f1-score was 100% and 93%, for Attack and The ROC curve of Random Forest model, PartOfAHorizontalPortScan, respectively.

We notice also that 'C&C-HeartBeat-FileDownload' is presented in this model, so the encoding process is changed starting from 0 to 9 according to the order that these labels appear in the classification report as shown in Fig. 5, Attack=0, C&C-HeartBeat-FileDownload=5, so on.



```
                            precision    recall  f1-score   support

                    Attack       0.99      1.00      1.00       826
                    Benign       0.95      0.57      0.71     39222
                       C&C       0.99      0.12      0.21      2997
          C&C-FileDownload       0.69      0.82      0.75        11
             C&C-HeartBeat       0.82      0.42      0.56        66
C&C-HeartBeat-FileDownload       1.00      0.50      0.67         2
                 C&C-Torii       1.00      0.50      0.67         4
                      DDoS       1.00      0.82      0.90     27921
              FileDownload       0.67      0.80      0.73         5
 PartOfAHorizontalPortScan       0.86      1.00      0.93    150359

                  accuracy                           0.89    221413
                 macro avg       0.90      0.65      0.71    221413
              weighted avg       0.90      0.89      0.88    221413
```

Fig. 5. The Classification Report of XGBoost Model.

The highest ROC curves area values were 100% for the Attack and 91% for both C&C-FileDownload and DDos labels as shown in Fig. 6.

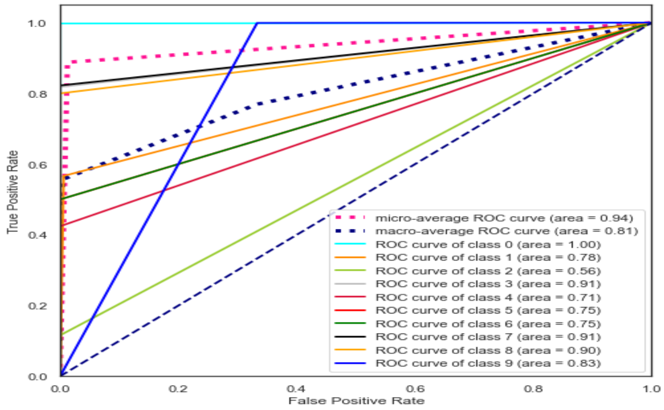
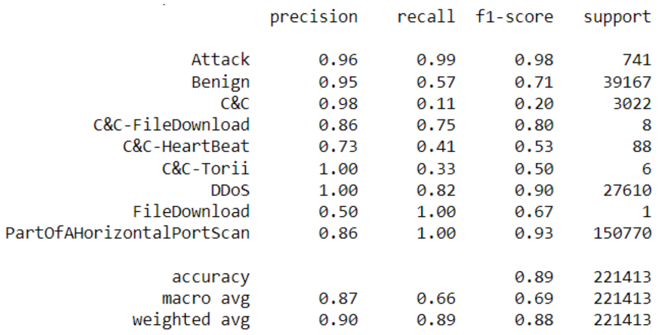


Fig. 6. The ROC Curve of XGBoost Model.

C. *CatBoost Classifier's Experiments*

CatBoost classifier reached the same overall accuracy as the Random Forest and XGBoost with a value of 89% as depicted in Fig. 7, but it takes more time than both of them. The 'C&C-HeartBeat-FileDownload' label does not appear in the classification report of this model since the very small number it has. The highest precision got is 100% for both C&C-Torii and DDoS, and also 96% for the Attack label. Furthermore, the highest recall was 100% for FileDownload and PartOfAHorizontalPortScan, 99% for the Attack, while the highest f1-score was 98%, 93%, 90%, for Attack, PartOfAHorizontalPortScan, and DDos labels, respectively.

On the other hand, the ROC curve was drawn for this model, where the highest area values were 100% for both Attack and FileDownload and 91% for the DDos label as shown in Fig. 8.



```
                           precision    recall  f1-score   support

                  Attack       0.96      0.99      0.98       741
                  Benign       0.95      0.57      0.71     39167
                     C&C       0.98      0.11      0.20      3022
         C&C-FileDownload       0.86      0.75      0.80         8
           C&C-HeartBeat       0.73      0.41      0.53        88
               C&C-Torii       1.00      0.33      0.50         6
                    DDoS       1.00      0.82      0.90     27610
            FileDownload       0.50      1.00      0.67         1
PartOfAHorizontalPortScan       0.86      1.00      0.93    150770

                accuracy                           0.89    221413
               macro avg       0.87      0.66      0.69    221413
            weighted avg       0.90      0.89      0.88    221413
```

Fig. 7. The Classification Report of CatBoost Model.

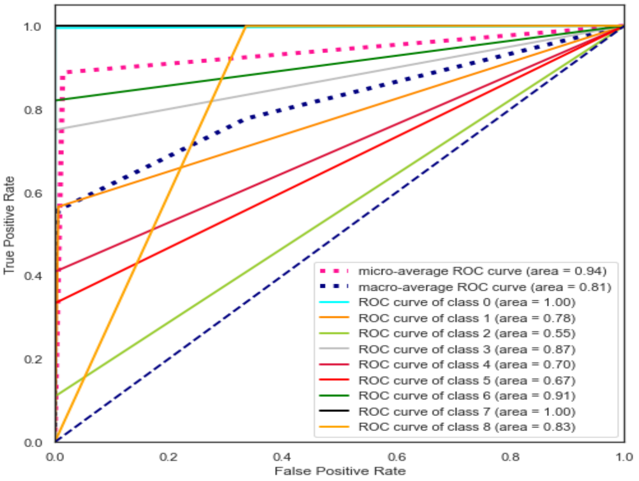
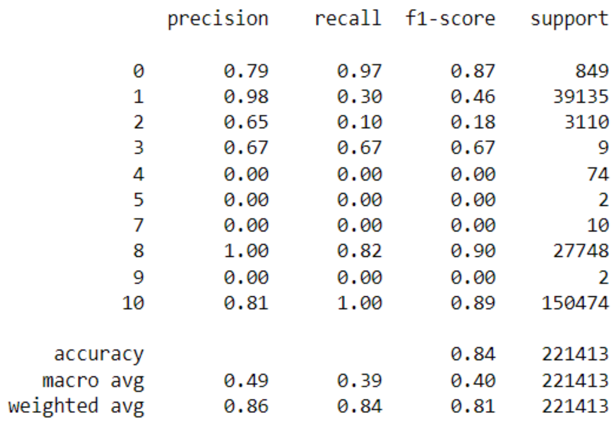


Fig. 8. The ROC Curve of CatBoost Model.

D. *CNN Classifier's Experiments*

CNN classifier is a deep learning algorithm is built based on many dense layers, as we explained in the methodology section. The proposed CNN model has ten layers and before train this model, the data is transferred into integers by applying 'MinMaxScaler' from the Sklearn library, and also the 'get_dummies' function from the Panda's library is applied to all used labels in this model, to change categorical variable into dummy/indicator variables. Unlike XGBoost, Random Forest, and CatBoost models, the classification report of the CNN model shows all used labels even if some labels have 0% for precision, recall, and f1-score metrics. However, the overall accuracy got by this model is lower than these models with a value of 84%, as shown in Fig. 9. The encoding numbers for this model is as follow: Attack=0, Benign=1, C&C=2, C&C-FileDownload=3, C&C-HeartBeat=4, C&C-HeartBeat-FileDownload=5, C&C-Mirai=6, C&C-Torii=7, DDoS=8, FileDownload=9, and PartOfAHorizontalPortScan=10. The C&C-Mirai label has not appeared in the classification report since it has only one case, while C&C-Torii, C&C-HeartBeat, C&C-HeartBeat-FileDownload, and FileDownload labels have appeared, but the measure matrices have 0%. The highest precision achieved is 100% for DDoS and 98% for Benign. Furthermore, the highest recall was 100% for PartOfAHorizontalPortScan, 97% for the Attack, while the highest f1-score was 90%, 89%, 87%, for DDos, PartOfAHorizontalPortScan, Attack labels, respectively.



```
              precision    recall  f1-score   support

           0       0.79      0.97      0.87       849
           1       0.98      0.30      0.46     39135
           2       0.65      0.10      0.18      3110
           3       0.67      0.67      0.67         9
           4       0.00      0.00      0.00        74
           5       0.00      0.00      0.00         2
           7       0.00      0.00      0.00        10
           8       1.00      0.82      0.90     27748
           9       0.00      0.00      0.00         2
          10       0.81      1.00      0.89    150474

    accuracy                           0.84    221413
   macro avg       0.49      0.39      0.40    221413
weighted avg       0.86      0.84      0.81    221413
```

Fig. 9. The Classification Report of CNN Model.

On the other hand, the ROC curve was drawn for this model, where the highest area values were 100% for FileDownload, C&C-FileDownload, Attack, and C&C-HeartBeat-FileDownload classes, 93% for DDos class as shown in Fig. 10. Moreover, the micro-average ROC curve achieved the highest area value with 98% as compared to the previous models, they only attained 94%. Furthermore, the macro-average ROC curve has a 'nan' value because of including the C&C-Mirai label data, when we remove this label from being trained, this value change to 82%, which is the highest value compared to the mentioned models.

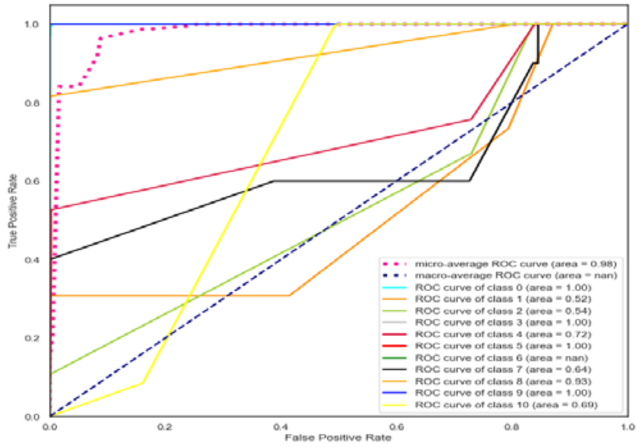
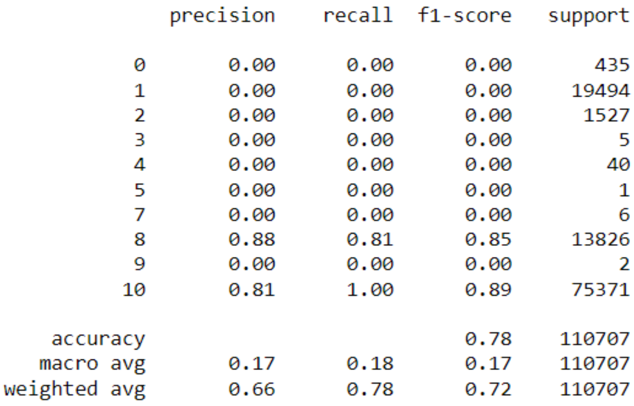


Fig. 10. The ROC Curve of CNN Model.

E. *LSTM Classifier's Experiments*

LSTM model is applied using various LSTM cells starting from 50 to 2000 cells, but the results were not good as the previous models. The highest overall accuracy is 78%; however, only DDoS and PartOfAHorizontalPortScan showed better results in the classification report of this model, while the rest had 0% for the used measure matrices, as depicted in Fig. 11. The encoding numbers for this model is exactly as in the CNN model.

On the other hand, the ROC curve was drawn for this model, where the highest area values were 99% for C&C-HeartBeat-FileDownload and C&C-FileDownload, 89% for C&C-Torii, 82% for DDoS, and 88% for PartOfAHorizontalPortScan label, as depicted in Fig. 12.



```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00       435
           1       0.00      0.00      0.00     19494
           2       0.00      0.00      0.00      1527
           3       0.00      0.00      0.00         5
           4       0.00      0.00      0.00        40
           5       0.00      0.00      0.00         1
           7       0.00      0.00      0.00         6
           8       0.88      0.81      0.85     13826
           9       0.00      0.00      0.00         2
          10       0.81      1.00      0.89     75371

    accuracy                           0.78    110707
   macro avg       0.17      0.18      0.17    110707
weighted avg       0.66      0.78      0.72    110707
```

Fig. 11. The Classification Report of LSTM Model.

These values are the lowest among all other used models. Furthermore, the micro-average ROC curve achieved an area value of 88% and the macro-average ROC area reached 57% when we removed class number 6 from being calculated.

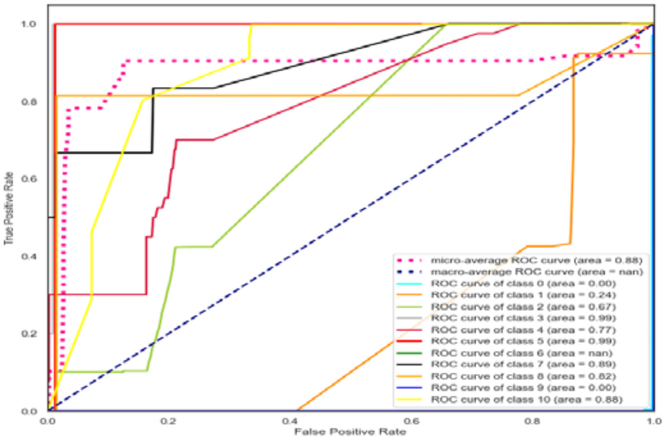
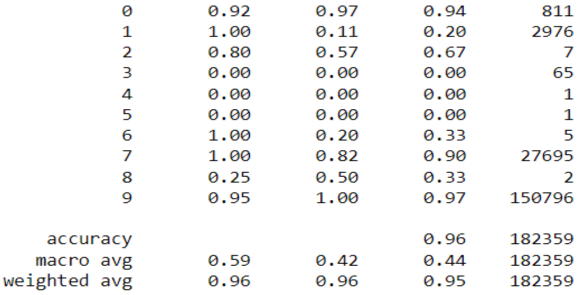


Fig. 12. The ROC Curve of LSTM Model.

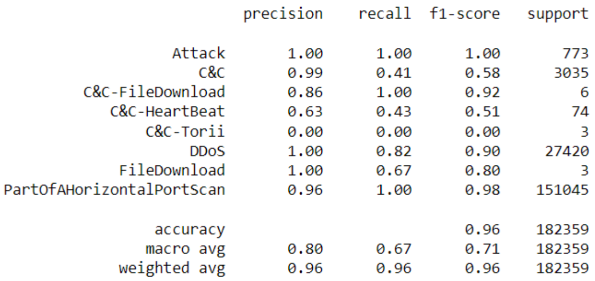
However, an experiment was conducted for all the classifiers used in this paper. First, by using the CNN model, the accuracy of this model reached 96% when the Okiru and benign labels were excluded from being trained and tested, as shown in Fig. 13.

Furthermore, the RF model when Okiru and benign labels were removed and achieved the same accuracy results as CNN, as depicted in Fig. 14.



```
           0       0.92      0.97      0.94       811
           1       1.00      0.11      0.20      2976
           2       0.80      0.57      0.67         7
           3       0.00      0.00      0.00        65
           4       0.00      0.00      0.00         1
           5       0.00      0.00      0.00         1
           6       1.00      0.20      0.33         5
           7       1.00      0.82      0.90     27695
           8       0.25      0.50      0.33         2
           9       0.95      1.00      0.97    150796

    accuracy                           0.96    182359
   macro avg       0.59      0.42      0.44    182359
weighted avg       0.96      0.96      0.95    182359
```

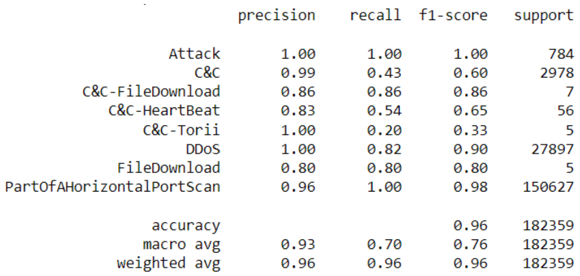
Fig. 13. The Classification Report of the CNN Model without Okiru and benign Labels.



```
                          precision    recall  f1-score   support

                  Attack       1.00      1.00      1.00       773
                     C&C       0.99      0.41      0.58      3035
         C&C-FileDownload       0.86      1.00      0.92         6
            C&C-HeartBeat       0.63      0.43      0.51        74
                C&C-Torii       0.00      0.00      0.00         3
                     DDoS       1.00      0.82      0.90     27420
             FileDownload       1.00      0.67      0.80         3
PartOfAHorizontalPortScan       0.96      1.00      0.98    151045

                 accuracy                           0.96    182359
                macro avg       0.80      0.67      0.71    182359
             weighted avg       0.96      0.96      0.96    182359
```

Fig. 14. The Classification Report of the RF Model without Okiru and benign Labels.

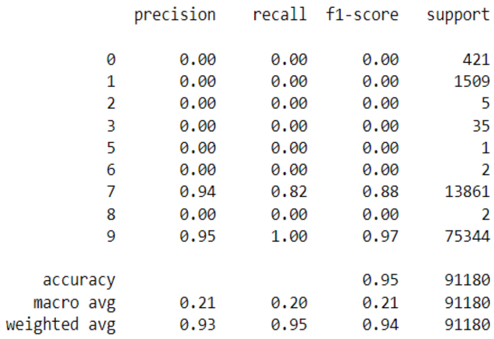
Additionally, by applying the same experiment as in CNN and RF model, the XGBoost model got the same accuracy as presented in Fig. 15.



```
                          precision    recall  f1-score   support

                  Attack       1.00      1.00      1.00       784
                     C&C       0.99      0.43      0.60      2978
         C&C-FileDownload       0.86      0.86      0.86         7
            C&C-HeartBeat       0.83      0.54      0.65        56
                C&C-Torii       1.00      0.20      0.33         5
                     DDoS       1.00      0.82      0.90     27897
             FileDownload       0.80      0.80      0.80         5
PartOfAHorizontalPortScan       0.96      1.00      0.98    150627

                 accuracy                           0.96    182359
                macro avg       0.93      0.70      0.76    182359
             weighted avg       0.96      0.96      0.96    182359
```

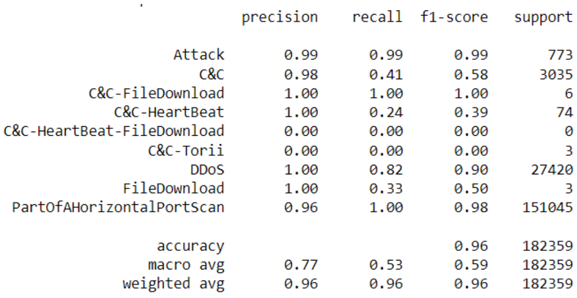
Fig. 15. The Classification Report of the XGBoost Model without Okiru and benign Labels.

Finally, the same accuracy was achieved too by using the Catboost classifier, as shown in Fig. 16, while the LSTM classifier got the lowest accuracy among all others, with a value of 95%, as depicted in Fig. 17.



```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00       421
           1       0.00      0.00      0.00      1509
           2       0.00      0.00      0.00         5
           3       0.00      0.00      0.00        35
           5       0.00      0.00      0.00         1
           6       0.00      0.00      0.00         2
           7       0.94      0.82      0.88     13861
           8       0.00      0.00      0.00         2
           9       0.95      1.00      0.97     75344

    accuracy                           0.95     91180
   macro avg       0.21      0.20      0.21     91180
weighted avg       0.93      0.95      0.94     91180
```

Fig. 16. The Classification Report of the Catboost Model without Okiru and benign Labels.



```
                           precision    recall  f1-score   support

                   Attack       0.99      0.99      0.99       773
                      C&C       0.98      0.41      0.58      3035
          C&C-FileDownload       1.00      1.00      1.00         6
             C&C-HeartBeat       1.00      0.24      0.39        74
C&C-HeartBeat-FileDownload       0.00      0.00      0.00         0
                 C&C-Torii       0.00      0.00      0.00         3
                      DDoS       1.00      0.82      0.90     27420
              FileDownload       1.00      0.33      0.50         3
 PartOfAHorizontalPortScan       0.96      1.00      0.98    151045

                  accuracy                           0.96    182359
                 macro avg       0.77      0.53      0.59    182359
              weighted avg       0.96      0.96      0.96    182359
```

Fig. 17. The Classification Report of the LSTM Model without Okiru and benign Labels.

V. Discussion

First, only 100000 rows from the twenty-malware traffic captures were loaded, in which 11 malware labels and one benign label have existed. We have done this way to use a balanced dataset that includes all labels. However, some steps must be carried out before using the classifiers, a preprocessing step was applied to remove irrelevant or missing data from the dataset, and then a feature engineering technique was performed to extract the most significant features and thus reduce the dataset's dimensionality. Three experiments have been done to investigate the best accuracy of detection based on multi-class classifications. The first one was to include all existed labels but the best accuracy achieved is not exceeded 74%. In the second and third experiments, one or more labels are removed from being classified. According to the experiments' outcomes done in the second one, RF, Catboost, and XGBoost models achieved the highest accuracy with 89% for all malware labels except the Okiru label. The CNN results were lower than RF, Catboost, and XGBoost algorithms, with a value of 84%. The worst accuracy was attained by using LSTM with a value of 78%, and we have to improve the parameters of cells size up to 4000 but not enhance. Furthermore, to enhance the accuracy more than the ones recorded in the first and second experiments, a third one was conducted. Furthermore, when we remove the benign and Okiru labels from being trained, the accuracy of RF, Catboost, and XGBoost algorithms raised to 96%. Besides, the CNN results recorded lower accuracy than RF, Catboost, and XGBoost algorithms, for all labels except Okiru; however, when we excluded the benign label data from being trained, the overall accuracy increased to 96% based on the classification report outcomes.

When comparing our methodology outcomes, almost the same accuracy of detection has been achieved as the authors had in [6] and [7], where the best-recorded accuracy was 74% for the best model by including all labels of the IoT-23 dataset. However, the model achieved a higher accuracy of detection reached 89% for the best model by including all labels except Okiru malware.

Moreover, some studies achieved a higher accuracy more than 96% which was attained by the best models of this work. However, these studies are not analyzed all existed labels of the IoT-23 dataset. For example, the authors in [8] involved detection for four malware types, while in [9] and [10] papers only five malware types were classified, and eight and nine malware types were analyzed in [11] and [12], respectively. Finally, in ML context RF and boosting algorithms are the best candidates for the proposed security system. Based on the performance of the RF in this paper it performs the training in less time. The authors [22] recommend Catboost for better prediction in their model. They have less consumed of time cost and high performance to embed in IoT devices to detect in real-time. Table III shows the comparison our models with other work.

TABLE III. Comparison our Models with other Work

Dataset	Number of labels	Technique	Accuracy
A. K. Sahu [13]	8	CNN	96 %
Dr. R. Thamaraiselvi [15]	binary	RF, Naive Bayes, Support Vector Machine, and Decision Tree	99 %
B. Roy et al. [23]	5	LSTM and BRNN	72 %
H. HaddadPajouh et al. [24]	CC	LSTM and BNN	84 %
Amiya Kumar Sahu[25]	9 without benign	CNN and LSTM	96 %
Our models all labels	12	RF, CatBoost, Xgboost, CNN, and LSTM	74 %
Our models without okiru	11	RF, CatBoost, Xgboost, CNN, and LSTM	89 % except CNN 84% and LSTM 78%
Our models without benign	10	RF, CatBoost, Xgboost, CNN, and LSTM	96 % except LSTM 95%

VI. Conclusion

The IoT environment deals with massive data that must be protected from being modified or stolen by attackers. This research paper proposed the application of five deep and ML algorithms to detect malware in network traffic based on the IoT-23 dataset. RF, Catboost, Convolutional Neural Network, Long Short-Term Memory (LSTM), and XGBoost classifiers are implemented and evaluated. Three experiments have been conducted for each classifier, to evaluate the best performance matrices that can be recorded among all classifiers. In the first evaluation task, the best accuracy achieved is 74% by using RF, Catboost, and XGBoost models for classifying all labels of the IoT-23 dataset. While the second experiment was to detect all labels except Okiru malware, the same models achieve the highest accuracy with a value of 89%. Finally, the last one was done without Okiru and benign labels, all models except LSTM get an accuracy of 96%.

In conclusion, machine and DL models can perform detection tasks with a high degree of accuracy and reliability. Analysis of a huge amount of traffic data can be accomplished to detect various types of intrusion. The paradigm of distributed detection can provide a deep analysis functionality particularly if the device operating pattern is involved. However, the available datasets in the literature are intended for central traffic characterization.

VII. Further Work

The work applied in this paper is limited to using malware traffic captures and only reading the first 100000 rows from each capture, which leads to unbalance data, especially for C&C-Mirai (1 one row), C&C-HeartBeat (341 rows), FileDownload (13), C&C-FileDownload (43), C&C-Torii (30), and C&C-HeartBeat-FileDownload (8). Therefore, to enhance the overall accuracy, we can remove the small size malware types from being classified. Besides, extra datasets can be used such as IoT Network Intrusion, BoT-IoT, and MQTT-IoT-IDS2020, to collect more balanced data for creating a big dataset to improve the detection accuracy, especially in DL techniques. Moreover, using more advanced CNN models such as vgg16 or vgg19 can be an avenue for further evaluation. Finally, the available datasets for evaluation consider centralized methodologies; the research community demands other datasets involving end-system operating patterns which may detect novel attack.

ACKNOWLEDGMENT

This research is supported by Aljouf University - Vice Presidency for Graduate Studies and Scientific Research as a part of scientific publication support program.

REFERENCES

[1] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, “Network intrusion detection for IoT security based on learning techniques,” IEEE Communications Surveys \& Tutorials, vol. 21, no. 3, pp. 2671–2701, 2019.

[2] T. Qiu, N. Chen, K. Li, M. Atiquzzaman, and W. Zhao, “Thirdquarter 2018,‘How can heterogeneous Internet of things build our future: a survey,’” IEEE Commun. Surv. Tutorials, vol. 20, no. 3, pp. 2011–2027, 2018.

[3] T. P. J. Kanimozhi V, “Artificial Intelligence for anomaly detection by employing deep learning strategies in IoT networks using the trendy IoT-23 big data from Google’s Tensorflow2.2,” Research Square, 2021, doi: https://doi.org/10.21203/rs.3.rs-364763/v1.

[4] “2022 SonicWall Cyber Threat Report,” sonicwall, 2022. https://www.sonicwall.com/2022-cyber-threat-report/.

[5] N.-A. Stoian, “Machine Learning for anomaly detection in IoT networks: Malware analysis on the IoT-23 data set,” University of Twente, 2020.

[6] N. Liang, Y. and Vankayalapati, “Machine Learning and Deep Learning Methods for Better Anomaly Detection in IoT-23 Dataset Cybersecurity,” Canada, 2021.

[7] SNEHA, “IoT Network Anomaly Detection Using Machine Learning and Deep Learning,” Rajasthan, INDIA, 2021.

[8] J. Vitorino, R. Andrade, I. Praça, O. Sousa, and E. Maia, “A Comparative Analysis of Machine Learning Techniques for IoT Intrusion Detection,” arXiv preprint arXiv:2111.13149, 2021.

[9] V. Dutta, M. Choraś Michałand Pawlicki, and R. Kozik, “A deep learning ensemble for network anomaly and cyber-attack detection,” Sensors, vol. 20, no. 16, p. 4583, 2020.

[10] 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.

[11] 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.

[12] N. Abdalgawad, A. Sajun, Y. Kaddoura, I. A. Zualkernan, and F. Aloul, “Generative Deep Learning to detect Cyberattacks for the IoT-23 Dataset,” IEEE Access, 2021.

[13] A. K. Sahu, S. Sharma, M. Tanveer, and R. Raja, “Internet of Things attack detection using hybrid Deep Learning Model,” Computer Communications, vol. 176, pp. 146–154, Aug. 2021, doi: 10.1016/j.comcom.2021.05.024.

[14] Bram. Dartel, “Malware detection in IoT devices using Machine Learning.,” University of Twente, 2021.

[15] Dr. R. Thamaraiselvi and S. Anitha Selva Mary, “Attack and Anomaly Detection in IoT Networks using Machine Learning,” International Journal of Computer Science and Mobile Computing, vol. 9, no. 10, pp. 95–103, Oct. 2020, doi: 10.47760/ijcsmc.2020.v09i10.012.

[16] D. Nanthiya, P. Keerthika, S. B. Gopal, S. B. Kayalvizhi, T. Raja, and R. S. Priya, “SVM Based DDoS Attack Detection in IoT Using Iot-23 Botnet Dataset,” in 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), Nov. 2021, pp. 1–7. doi: 10.1109/i-PACT52855.2021.9696569.

[17] B. M. Khammas, “The Performance of IoT Malware Detection Technique Using Feature Selection and Feature Reduction in Fog Layer,” IOP Conference Series: Materials Science and Engineering, vol. 928, no. 2, p. 022047, Nov. 2020, doi: 10.1088/1757-899X/928/2/022047.

[18] T. G. Palla and S. Tayeb, “Intelligent Mirai Malware Detection for IoT Nodes,” Electronics (Basel), vol. 10, no. 11, p. 1241, May 2021, doi: 10.3390/electronics10111241.

[19] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, and S. Venkatraman, “Robust Intelligent Malware Detection Using Deep Learning,” IEEE Access, vol. 7, pp. 46717–46738, 2019, doi: 10.1109/ACCESS.2019.2906934.

[20] S. Garcia, A. Parmisano, and M. J. Erquiaga, “IoT-23: A labeled dataset with malicious and benign IoT network traffic (Version 1.0. 0)[Data set]. Zenodo.” 2020.

[21] N. Abdalgawad, A. Sajun, Y. Kaddoura, I. A. Zualkernan, and F. Aloul, “Generative Deep Learning to detect Cyberattacks for the IoT-23 Dataset,” IEEE Access, 2021.

[22] C. Bentéjac, A. Csörg\Ho, and G. Mart\’\inez-Muñoz, “A comparative analysis of gradient boosting algorithms,” Artificial Intelligence Review, vol. 54, no. 3, pp. 1937–1967, 2021.

[23] B. Roy and H. Cheung, “A Deep Learning Approach for Intrusion Detection in Internet of Things using Bi-Directional Long Short-Term Memory Recurrent Neural Network,” in 2018 28th International Telecommunication Networks and Applications Conference (ITNAC), Nov. 2018, pp. 1–6. doi: 10.1109/ATNAC.2018.8615294.

[24] H. HaddadPajouh, A. Dehghantanha, R. Khayami, and K.-K. R. Choo, “A deep Recurrent Neural Network based approach for Internet of Things malware threat hunting,” Future Generation Computer Systems, vol. 85, pp. 88–96, Aug. 2018, doi: 10.1016/j.future.2018.03.007.

[25] A. K. Sahu, S. Sharma, M. Tanveer, and R. Raja, “Internet of Things attack detection using hybrid Deep Learning Model,” Computer Communications, vol. 176, pp. 146–154, Aug. 2021, doi: 10.1016/j.comcom.2021.05.024.

Power Grid Resource Integration of Enterprise Middle Station based on Analytic Hierarchy Process

Shaobo Liu, Li Chen, Wenyuan Bai, Zhen Zhang, Fupeng Li*
State Grid Gansu Electric Power Company
Lanzhou Gansu730030, China

***Abstract*—With the transformation from smart grid to power Internet of Things, new power businesses such as power grid automation and power quality monitoring are constantly emerging. The load environment of power grid is changeable. In order to meet the needs of multi-service, the integrated access scheme for power grid resources in power enterprises is gradually diversified, which brings challenges to the unified management and control of power grid communication network. In this paper, SDN technology is used to improve the operation and maintenance management and control of power communication network, which aims at the integration scheme of power grid resources in power enterprises. Based on the controller cluster technology, combined with the new power business requirements, this paper designs a software-defined network centralized control architecture of the new business of power communication network. The architecture realizes the operation and maintenance management of network resources under the centralized control architecture of typical enterprise scenarios, such as power grid enterprises. The convergence speed is improved by 27%. The minimum value of iterative convergence is 31% better than that of other methods. The system requirement is reduced by 13.5%, which is helpful to improve the efficiency of node dynamic allocation and ensure the need of large-capacity data transmission of smart grid. The research in this paper can realize the two-way interaction, real-time expansion and unified deployment of power business in the future, and promote the intensive and lean development of power communication network.**

Keywords—Power grid enterprises; SDN; power grid resources; centralized control architecture; power communication network

I. Introduction

Under the environment of orderly release of new policies, such as power generation, consumption plan and incremental business of power grid, all kinds of users will actively participate in the related industries of power distribution and consumption side. At the same time, with the development of sensing, edge computing, big data processing and other technologies, the transformation from smart grid to power Internet of Things is accelerating. The information exchanges with things and things, people and things is more frequent [1]. The power grid resource business center cooperates with the customer service center. The data center and the IOT management center to jointly support the power grid business applications, such as power outage analysis, accurate fault research and judgment, and line loss analysis in the same period [2]. It can arrange enterprise-level business services to solve the problems of non-standard business services, difficult precipitation and insufficient sharing.

Resource aggregation in China and abroad has also been widely concerned and discussed by experts and scholars. The author in [3] uses the panoramic theory to integrate a variety of micro grid resources, which takes into account the complementarity of various distributed generations. micro grid aggregation not only minimizes the damage to the power grid caused by the output fluctuation of the micro grid, but also improves the security of the system and ensures the optimization of its own economic benefits. The author in [4] considers the uncertainty of EV behavior characteristics of Load Aggregator (LA), and uses this method to integrate discrete EV energy storage. The author in [5] puts forward the resource evaluation system and resource quality distinction criteria for demand response, and builds a demand response resource integration model for multiple user load groups. In [6], Virtual Power Plant (VPP) is introduced into the wind power and EV, with power balance as the constraint. Taking the highest comprehensive economic benefit of its discharge as the goal, the mathematical model of the power generation plan is constructed.

The research proposes a scheme for fast recovery of the power communication network after a fault occurs. To make the time delay of the power business meet the provisions of the standard IEC 61850, a new network calculus method is proposed. Author proposes an SDN platform, which is suitable for the large-scale intelligent meter reading system, and can balance the load of power communication network through flow monitoring, load assessment, active control and path selection. However, with the gradual expansion of the distribution communication network, there are some problems such as low efficiency of network management and control, lack of rationality of the architecture, and lack of reliability. The research proposes an algorithm, which can realize the load scheduling of the QoS. Meanwhile, in order to reduce the delay of service transmission, the algorithm adopts the scheme that the multiple SDN controllers cooperate to manage the traffic of power communication network, but there is still the problem of complex protocol conversion between different communication media. The research innovatively introduces the SDN centralized control architecture into the power wide area communication network, and proposes a new fair resource allocation algorithm for power services, which is used to ensure that high-priority services are allocated to the shortest path for transmission and maximize throughput. Considering the burden of communication network, it affects the operation and maintenance management and control of power communication network.

The main innovations of this paper are:

1) This paper proposes to use the NFV technology to complete the software of various terminal hardware functions of general hardware.

2) Improve all kinds of power distribution and consumption monitoring and control equipment, centralize the proprietary hardware into a general hardware, and use software to realize its functions.

3) Based on the analytic hierarchy process (AHP), a multi-objective decision-making algorithm is proposed to the multi-domain communication network of power business, which takes the performance of the enterprise middle station network as the judgment index.

II. Related Work

A. The Overall Structure of the Enterprise

In the environment of power Internet of Things, the information transmission between power users and power grid becomes more frequent. The demand for power consumption quality, real-time and reliability of information is higher and higher. The safe and efficient power communication network is the key link to cope with this demand. The comprehensive control function of the power communication network is the most important one [7]. This paper is based on the architecture of the resource control system of enterprise middle station as shown in Fig. 1.

The architecture covers three parts: power business data forwarding layer, SDN cluster control system and software-based power business layer. The first layer mainly includes passive optical network, wireless private network, industrial Ethernet and other hardware devices. The NFV technology is used to transform the switches and other equipment of the multi-domain accesses network. The OpenFlow general protocol is used to realize the integrated management and control of the cluster control system over the underlying equipment routing, bandwidth allocation, inter-domain switching and other functions. The underlying switch only needs to complete simple information forwarding and flow table matching tasks [8]. The southbound interface, which is responsible for the communication with the terminal, is responsible for providing a channel for information interaction with the control system. At the same time, distribution transformer videos surveillance, power consumption information acquisition and other services are connected to the data forwarding layer through the underlying hardware devices so that the control system can flexibly monitor and control them.

B. Power Communication Network SDN Operation and Maintenance Control

Realize the centralized operation, maintenance, management and control of power communication network, optimize the allocation of resources, and increase the flexibility and compatibility of the network. The idea of applying the SDN to the current power communication networks operation and maintenance management is shown in Fig. 2.

This paper uses the method of operation and maintenance management and control for reference to improve various power distribution monitoring and management and control devices. It concentrates the proprietary hardware on a general hardware, uses software to realize its function, and makes it support programmable network interface.

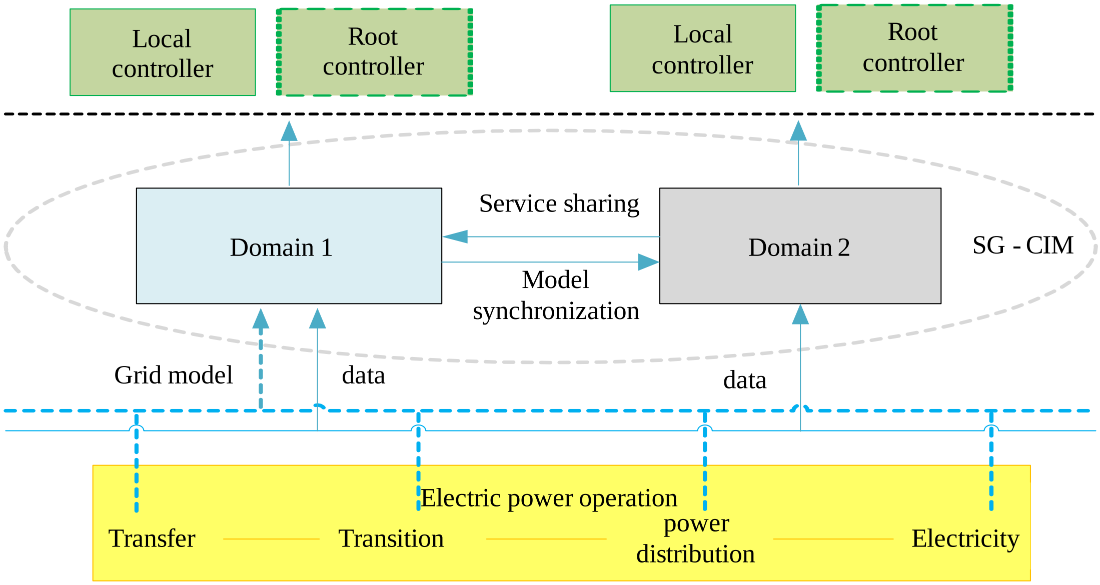


Fig. 1. Enterprise Mid-stage Control Architecture.

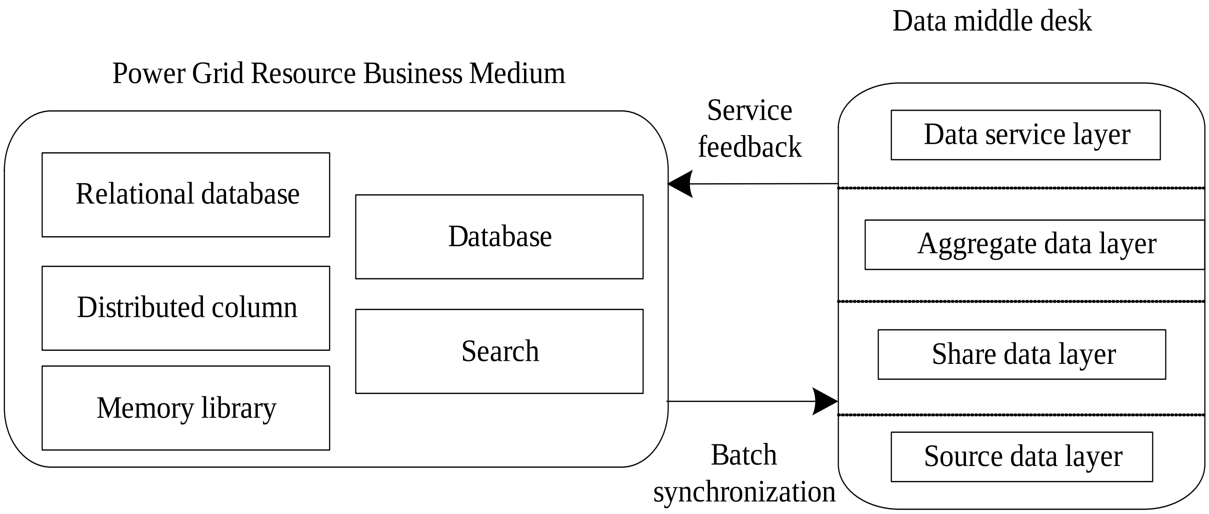


Fig. 2. Network Operation and Maintenance Control based on SDN Technology.

III. SDN High Reliability Control Scheme based on Cluster Technology

In the SDN control architecture, the controller is in the core position and is the basis of realizing the control function. How to deploy the controller and how to achieve the function is the key to realize the centralized management and control of the power communication network. This paper chooses the relatively mature and perfect controller cluster technology in the industry to complete the flexible deployment of various controllers in the architecture control layer, and realizes the operation and maintenance management and control functions of the power communication network through software [9]. The network realizes the data interaction between different functional controllers through a unique cluster communication process. The SDN control architecture based on the controller cluster is shown in Fig. 3.

The architecture is divided into three layers from top to bottom. The first layer is the cluster management layer, which mainly completes the functional task allocation of the root controller and the information synchronization between them. The root controller layer is responsible for ensuring the consistency of the overall state of the network and processing the corresponding business in time. The third layer is the local controller layer, which is responsible for handling complex and diverse local services. At the bottom are some switches, which are used to forward and process the underlying information [10].

The application of cluster technology to the control system can make the centralized control architecture reduce the risk of single point failure of control equipment under the original control mode, which increases the good scalability. It also copes with tens of thousands of switch traffic, significantly shorten the transmission delay of southbound interface protocol packets, and optimize the network transmission quality [11]. For any network G, there are three parts, namely network node devices, communication links, and topology [12]. The reliability R of the network G mainly depends on the reliability L_n of the node equipment. The reliability of the link L_1 and the topology N.’ s Functional relationship is as follows:

$$R \in \left(L_n, L_1, N\right) \tag{1}$$

Formula 1 shows that the network reliability is directly proportional to the reliability quality of the three network factors.

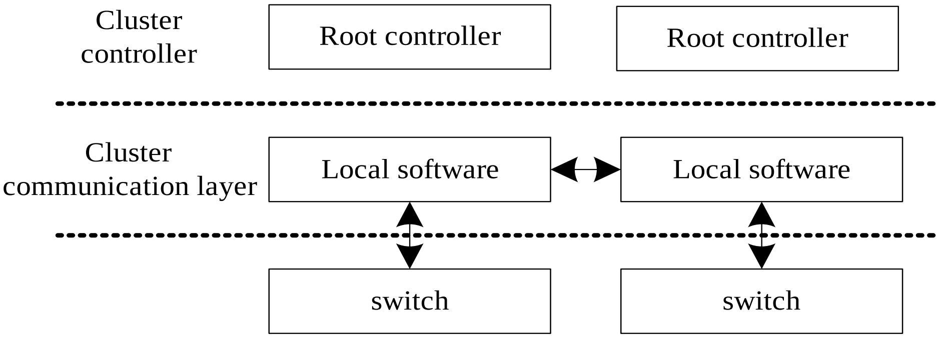


Fig. 3. SDN Control Scheme based on Cluster Technology.

IV. Interdomain Switching Algorithm based on the Analytic Hierarchy Process

When choosing the algorithm, we need to use the appropriate algorithm according to the actual needs. In the evaluation of multi-type resource characteristics of the multiple users participating in the aggregation by the analytic hierarchy process (AHP), the dimension of the clustering vector is low, so this paper analyzes the inter-domain hierarchical switching.

The analytic hierarchy process (AHP) analyzes various influencing factors of things from different perspectives. Its main steps are shown in Fig. 4.

1) Construct multi-index decision matrix: Firstly, on the basis of analyzing the target of the problem, the scheme set *X* to be selected and the index matrix *S* to make judgment are obtained.

Where $X=\{x_1,x_2,\cdots,x_m\}$ represents that there are *m* alternatives. $S=\{s_1,s_2,\cdots,s_n\}$ indicates that there are *n* evaluation indicators. Then, a multi-index evaluation matrix is constructed:

$$A=(\alpha_{ij})_{mn}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{2}$$

Where a_{ij} represents the result of alternative plan x_i relative to the judgment index s_i.

2) Vector normalization method

The matrix A is normalized.

$$b_n=\frac{a_n}{\sqrt{a^2 a_{ij}^2}} \ ,i=1,2,\cdots,m \tag{3}$$

Where, matrix b is the benefit matrix, matrix a is the allocation matrix, i is the set of nodes accessible to ants other than the elements in the tabu list, and a_n represents whether to allocate a frequency band to n users.

3) The characteristics of different parameters are analyzed and compared to establish the corresponding comparison matrix.

Compare each parameter of the candidate scheme. Take the relative importance of the impact factors as the standard. Conduct pairwise comparison. Finally, evaluate different levels according to Table I [13].

4) Perform pairwise comparison on the parameters of different levels, wherein the comparison result forms a judgment matrix $C=\left(c_{ij}\right)$ As shown in the formula (4), the parameter *i* in the matrix $c_{ij}>0$ is more important than the parameter *j*.

$$C=(c_{ij})_{n\times n}=\begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \cdots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix} \tag{4}$$

For the comparative decision matrix C:

$$CI=\frac{\lambda_{max}(C)-N}{RI} \tag{5}$$

Where CI represents the consistency index, RI represents the average random consistency index, and CR represents the random consistency ratio, and $\lambda_{\max}$ is the increment of the solution converted into pheromone.

$$CR=\frac{CI}{RI} \tag{6}$$

Where *n* denotes the order of the comparison matrix. $\lambda_{\max}\left(c\right)$ represents the largest eigenvalue of the comparison matrix C[14].

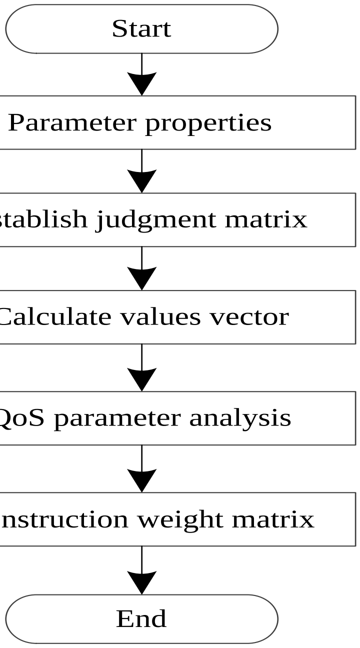


Fig. 4. Power Resource Hierarchical Analysis Process.

TABLE I. A Comparison Table of the Importance of Different Parameters

Comparison of the importance of *i* and *j*	The value of a_{ij}
Equal	1
A little important	3
Important	5
Very important	7
Very strong importance	9

When CR < 0.1, it indicates that the consistency check of the comparison matrix C constructed by AHP is qualified. When CR > 1, it indicates that the consistency check of the comparison matrix C constructed by AHP is unqualified. At this time, the judgment matrix A needs to be reconstructed until the consistency check is qualified [15].

5) The calculation judgment index is the weight of each factor: Assume that the characteristic vector corresponding to the maximum eigenvalue $\lambda_{\max}(C)$ of the comparison matrix C of the decision index is shown in Table II.

According to the comparison matrix of different candidate schemes under each index, the respective weight matrix is calculated as follows:

$$W'' = \{w_1', w_2'' \cdots w_n'\} \tag{7}$$

$$W^P = \{w_1^p, w_2^p \cdots w_n^p\} \tag{8}$$

Finally, the distribution coefficients of the different alternatives can be obtained by multiplying W^P and W'. For weight sorting, the scheme corresponding to the maximum weight is the best scheme at the total level. The decision function is expressed as:

$$f_{AHP} = \max\left(W^p \times W'\right) \tag{9}$$

The flow chart of the decision algorithm is shown in Fig. 5.

TABLE II. THE RI VALUE OF THE ORDER COMPARISON MATRIX

m	RI
1	0.00
2	0.00
3	0.58
4	0.90
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45
10	1.49

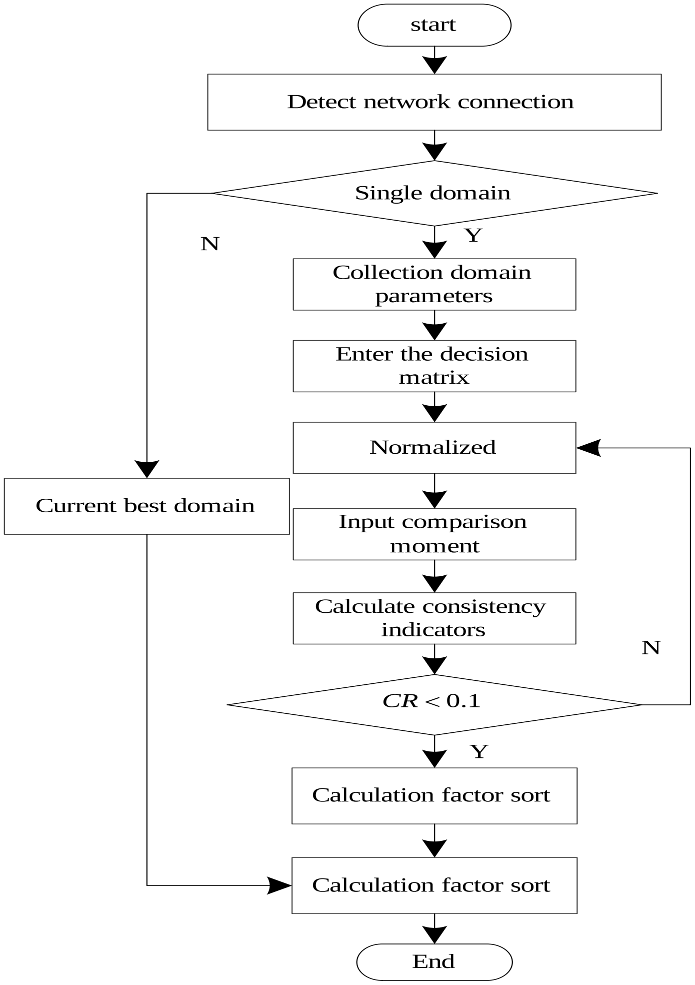


Fig. 5. AHP Algorithm Flow Chart.

V. Simulation Experiment Analysis

A. Contrast Solution

In order to verify the performance of the algorithm in this paper, the following algorithm is used as a comparison scheme.

1) RAR _ RS _ BiG: according to the algorithm idea, first perform augmentation operation on each SFC (Service Function Chain), that is, add backup VNF (Virtualised Network Function） in each SFC chain to meet the availability of the whole chain. Where: RAR is the file attribute.

2) JDBS _ s: all VNF instances in the JDBS _ s algorithm are implemented based on the single-tenancy technology.

3) GREP: The core idea of GREP to improve the availability of AHP is to provide backup for the VNF with the lowest availability in its corresponding SFC. Then, gradually provide backup for the less reliable VNF until the availability requirements of SFC are met.

B. Parameter Settings

For AHP, the number of VNFRs contained in each AHP is an integer randomly selected from the list [3, 4, 5, 6]. The CPUDRC, memory DRC and bandwidth resource requirements of each VNFR follow the uniform distribution of (10, 50) units. There are 10 kinds of VNFs in the simulation. The initial availability of each original VNF and the availability of each backup VNF are subject to the uniform distribution of (0.99, 0.999). The number of AHP to be mapped, the availability requirements of AHP, and the CPU BRC and memory BRC required to instantiate a VNF will be used as variables in different simulation scenarios

C. Simulation Results and Performance Analysis

1) Comparative analysis of processing ability: In order to reduce the impact on errors, each set of results is based on the statistical values obtained from 10 experiments. The error bar to the figure represents the 95% confidence interval. In the simulation, enough resources are set for the data center so that all AHPs can be mapped to the network. Therefore, in each set of experimental results, the final number of servers started, the amount of CPU and memory backup resources consumed and the BRC can be used as indicators to measure the performance of each algorithm. The final performance comparison between the changed AHP availability Fig. 6 shows the resource consumption comparison and the final BRC comparison of each algorithm under the changed AHP availability.

In the experiment of this scenario, the number of AHP to be mapped is 1000. The CPU BRC and memory BRC required for instantiating a VNF are both set to 20 units.

As shown in Fig. 4.2(a), it is a comparison of the final number of servers used by each algorithm. From the results, it can be seen that the number of servers used by the algorithm in this paper is the least. Compared with RAR _ RS _ BiG, JDBS _ s and GREP, they can use up to 29%, 29% and 42% of the servers respectively. JDBS _ s is the single-tenancy version of the roBS algorithm, that is, except for the difference in the implementation of VNF and IJ, other algorithms, such as AHP mapping algorithm, resource reservation algorithm, and VNFR merging algorithm of the same type, are the same as the method in this paper. By comparing the algorithms, it can be seen that when the availability requirement of AHP is lower than the initial availability of VNF, that is, the experimental scenarios with availability requirements of 0.90 and 0.99, the two algorithms have similar performance. This is because in this case, the availability of the current VNF can meet the needs of AHP. There is no need to reserve a backup for each original VNF.

2) Comparative analysis of algorithm performance: In this paper, based on the AHP algorithm, Matlab2014b is used as the simulation software to select the multi-domain network for the integrated automatic business of power grid resources in the enterprise middle station. The selection results are recorded for comparative analysis.

The electric power communication system requires high network security. There are many video surveillance services, which are interactive. The bandwidth is the key to ensure high-quality video streaming, so it requires a large communication bandwidth and a relatively high transmission delay. The demand for packet loss rate is weaker than the former two. The comparison matrix is shown in Table III.

The maximum eigenvalue $\lambda_{\max}\left(C\right)$ and eigenvector γ of the matrix C are obtained by using the eig function in MATLAB. It can be concluded that the CR value is less than 0.1 through formulas (10) and (11). The decision matrix meets the consistency requirements and can be used to calculate the weight vector. The weight vector is normalized to obtain the single factor and weight coefficient combination of service 1:

$$W^{'} = \{w_1, w_2, w_3, w_4\} = \{0.0469, 0.0926, 0.3641, 0.4964\} \tag{10}$$

$$W^{P} = \begin{pmatrix} w_1, w_2, w_3, w_4 \\ w_1', w_2', w_3', w_4' \end{pmatrix} \tag{11}$$

According to the decision function, the distribution coefficient of the total hierarchy is calculated. The simulation results are shown in Fig. 7.

The results are:

$$ts = \left(0.2715, 0.7285\right) \; f_1 = \max\left(ts\right) = 0.7345$$

Where 0.7285 represents the utility value f_1 of the best access domain 2 of the service, i.e., the distribution transformer video surveillance service in the current environment. Similarly, 0.2715 represents the utility value f_1 of the best access domain 1 of the service in the current environment. For the service f_1, domain 2 shall be selected as the best handover target network.

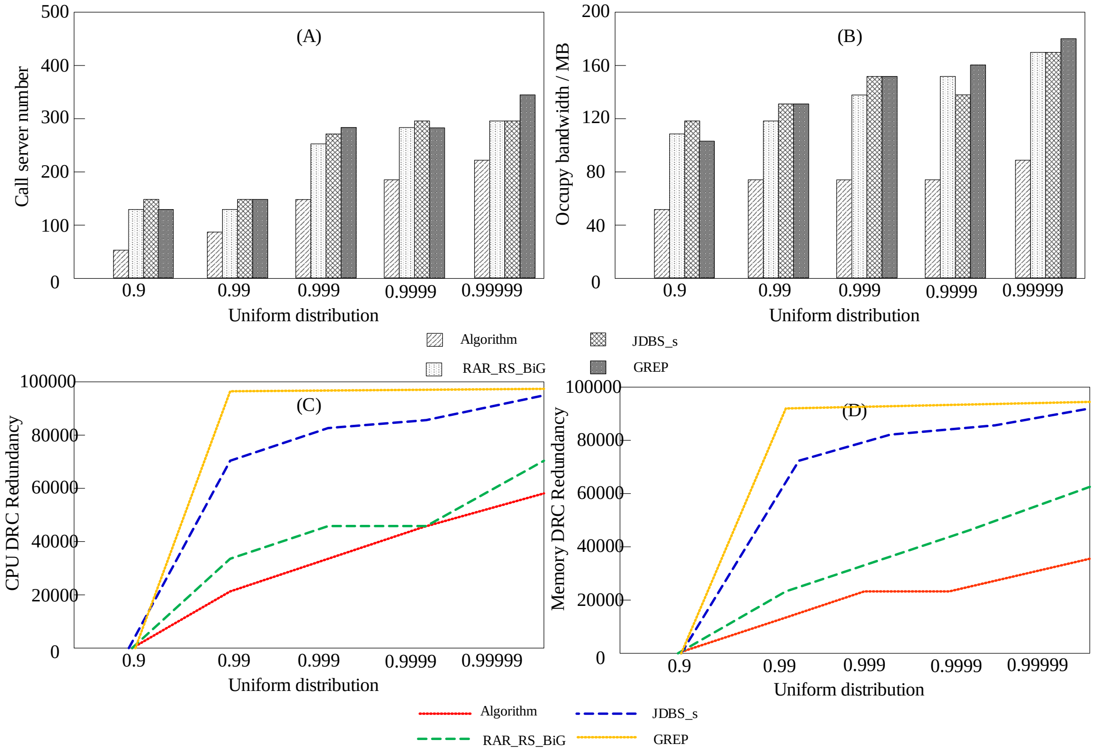


Fig. 6. Resource Consumption Comparison of each Algorithm and the Final BRC Comparison.

TABLE III. DISTRIBUTION VIDEO SURVEILLANCE SERVICE COMPARISON MATRIX

Comparative indicators	Time delay	Bandwidth	Packet loss rate	Hop count
Time delay	1	1/3	5	5
Bandwidth	3	1	8	9
Packet loss rate	1/5	1/8	1	2
Hop count	1/5	1/9	1/2	1

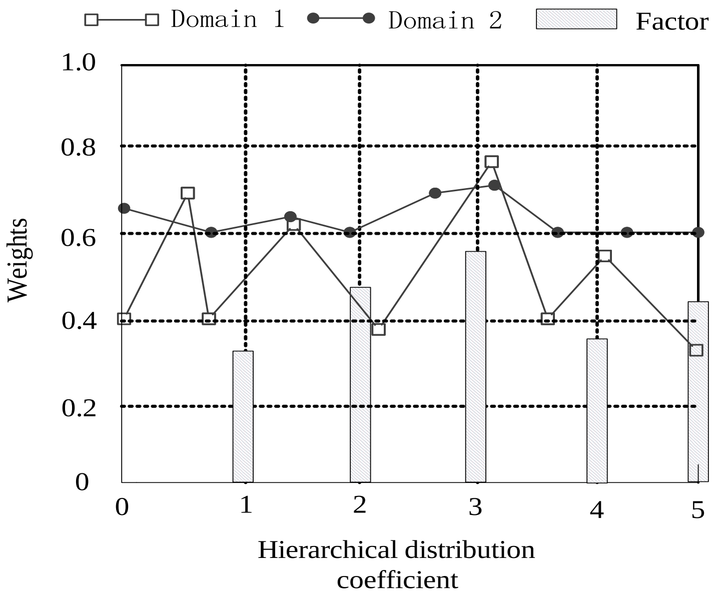


Fig. 7. Simulation Results.

In the aspect of electric power business, this paper selects typical business to study. In view of the intelligent electric power business, the analysis of its demand for the electric power communication network is more conducive to practical application. In addition to the control system, the intelligent electric power equipment is also an important component of the electric power communication network. Based on the analysis of the overall control system, this paper carries out a more in-depth and comprehensive study of the control system of the electric communication network.

VI. Conclusion

In this paper, the key technologies of the future electric power data Wan communication are studied. Through the research of network intelligent operation and maintenance and business scheduling strategy in enterprise platform environment, the research of resource elastic configuration and integration technology for IP network and optical transmission network, the efficient use of the IP layer and the optical layer network resources is realized, and the reliable operation of business is ensured, which is beneficial to that access of new equipment of a subsequent network and the rapid open of new services. Main work is as follows:

1) Design the overall framework of software defined network centralized control for the new business of electric power communication network.

2) The realization of network resource operation, maintenance management and control under centralized control architecture combined with typical application scenarios such as virtual power plant is done.

3) The simulation software is used to verify the realization of typical functions of SDN centralized control system.

The intelligent power equipment is also an important component of the electric power communication network. This paper is only based on the overall control system, and does not conduct in-depth research and design of intelligent power equipment. In the future, the optimization objectives and constraints in the construction of resource aggregates can be further improved, for example, the aggregation degree of resources can be optimized when considering spatial differences to achieve the simultaneous optimization of the characteristics and regions of resource aggregates.

References

[1] Liu Baoju, Yu Peng, Feng Lei, Qiu Xuesong, Jiang Hao. Power SDN communication network facing load balancing routing restructuring [J]. Journal of Beijing Post and Telecommunications University, 2020,43 (02): 16-22.

[2] Ding Huixia, Gao Kaiqiang, Zhang Geng, Wang Yanan. Research and Design of Power Communication Network channel simulation system [J]. Electronic devices, 2020,43 (05): 1056-1060.

[3] Zhao Hongda, Wang Zhe, Zhu Mingxia, Wang Haiyong. 5G communication technology in the application of Pan power IoT [J]. Southern Power Grid Technology, 2020,14 (08): 9-17.

[4] Zhang Jie, Xun Xiaoshuo, Xu Qi camphor. The application of ultra-broadband communication in the power system [J]. Electronic device, 2020,43 (03): 558-562+606.

[5] N. Dorsch, F. Kurtz, C. Wietfeld, “Enabling Hard Service Guarantees in Software-Defined Smart Grid Infrastructures,” Computer Networks, 2018, 147(01), 112-131.

[6] B. A. Maroua, K. K. Nguyen, C. Mohamed, “QoS-aware software-defined routing in smart community network,” Computer Networks, 2018, 147(01), 221- 235.

[7] P. A. Ribeiro, L. Duoba, R. Prior, “Real-time wireless data plane for real-time-enabled SDN. 15th IEEE International Workshop on Factory Communication Systems (WFCS),” Sundsvall, Sweden: IEEE Press, 2019, pp. 1-4.

[8] Z. Liu, S. Wang, Y. Liu, “Secrecy Transmission for Femtocell Networks Against External Eavesdropper,” IEEE Transactions on Wireless Communications, 2018, 17(8), pp. 5016-5028.

[9] S. Xiao, X. Zhou, Y. Yuan-Wu, “Robust Resource Allocation in Full-Duplex-Enabled OFDMA Femtocell Networks,” IEEE transactions on wireless communications, 2017, 16(10), pp. 6382-6394.

[10] M. Chiang, C. W. Tan, D. P. Palomar, “Power Control By Geometric Programming,” IEEE Transactions on Wireless Communications, 2017, 6(7), pp. 2640-2651.

[11] Z. Liu, Y. Yuan, H. Yuan, “Power Allocation Based on Proportional-Integral Controller inFemtocell Networks With Consideration of Maximum Power Constraint,” IEEE Systems Journal, 2018, 13(1), pp. 88-97.

[12] Z. Liu, S. Li, K. Ma, “Robust Power Allocation Based on Hierarchical Game with Consideration of Different User Requirements in Two-tier Femtocell Networks,” Computer Networks, 2017, 122, pp. 179-190.

[13] H. Haci, H. Zhu, J. Wang, “Performance of Non-orthogonal Multiple Access With a Novel Asynchronous Interference Cancellation Technique,” IEEE Transactions on Communications, 2017, 65(3), pp. 1319-1335.

[14] Y. Fu, Y. Chen, C. W. Sung, “Distributed Power Control for the Downlink of Multi-cell NOMA Systems,” IEEE Transactions on Wireless Communications, 2017, 16(9), pp. 6207-6220.

[15] V. Ozduran, E. Soleimani-Nasab, B. S. Yarman, “Effects of Co-channel Interference on Sum-rate Based Relay Selection Method for a Dual-hop Multiple Full-Duplex Two-way Wireless Relaying Networks,” Istanbul University Journal of Electrical & Electronics Engineering, 2017, 17(2), pp. 3387-3397.

A Screening System for COVID-19 Severity using Machine Learning

Abang Mohd Irham Amiruddin Yusuf [1], Marshima Mohd Rosli[2], Nor Shahida Mohamad Yusop[3]
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450, Selangor, Malaysia[1, 2, 3]
Institute for Pathology, Laboratory and Forensic Medicine, University Teknologi MARA[2]
47000 UiTM, Sungai Buloh, Selangor, Malaysia[2]

***Abstract*—COVID-19 disease can be classified into various stages depending on the severity of the patient. Patients in severe stages of COVID-19 need immediate treatment and should be placed in a medical-ready environment because they are at high risk of death. Thus, hospitals need a fast and efficient method to screen large numbers of patients. The enormous amount of medical data in public repositories allows researchers to gain information and predict possible outcomes. In this study, we use a publicly available dataset from Springer Nature repository to discuss the performance of three machine learning techniques for prediction of severity of COVID-19: Random Forest (RF), Naïve Bayes (NB) and Gradient Boosting (GB). These techniques were selected for their good performance in medical predictive analytics. We measured the performance of the machine learning techniques using six measurements (accuracy, precision, recall, F1-score, sensitivity and specificity) in predicting COVID-19 severity. We found that RF generates the highest performance score, which is 78.4, compared with NB and GB. We also conducted experiments with RF to establish the critical symptoms in predicting COVID-19 severity, and the findings suggested that seven symptoms are substantial. Overall, the performance of various machine learning techniques to predict severity of COVID-19 using electronic health records indicates that machine learning can be successfully applied to determine specific treatment and effective triage.**

Keywords—Severity prediction; COVID-19; random forest; Naïve Bayes; gradient boosting

I. Introduction

Research has shown that electronic health records (EHR) are becoming increasingly valuable to predict health outcomes or disease diagnoses [1]. For many years, researchers have been analysing EHR with statistical and machine learning techniques for prognostic evaluations. Statistical techniques are designed to determine relationships between variables, and machine learning techniques are designed to make the most accurate predictions based on EHR. Although both techniques play an important role in research, previous analysis has shown that machine learning outperforms statistical techniques because of the recent advancement of tools for data analytics and large quantity of data humanity has access to since the information explosion [2]–[4].

A significant number of prediction models utilising EHR have been proposed in the literature over past years, including in the recent COVID-19 pandemic [1]. The COVID-19 pandemic presented massive data for examining social, behavioural, public health, and economic impacts [5]. Many recent studies have conducted data mining analysis using various available datasets that focus on the occurrence of confirmed, fatal and recovered COVID-19 cases globally to understand the threats and predict the subsequent planning of containment activities [6]–[9].

The severity of COVID-19 disease varies from mild to critical stages. Screening of positive COVID-19 patients at primary care clinics is used as the initial triage to determine the severity stage and admission to hospital [10]. With use of the conventional methods, the process to manage COVID-19 patients in the initial triage is not efficient because of long waiting times for screening [11]. It has previously been observed that limited hospital resources and staff during the COVID-19 pandemic makes it important to decide which patients require more urgent treatment and which patients that can wait [12]. Therefore, a timely clinical decision is important to help in early detection of serious disease and provide effective treatment for the individual patient, which is important for reasonable allocation of medical resources.

In recent years, machine learning techniques have attained popularity in the health area because of their capability to deal with enormous, complex and unbalanced data, and yield outcomes such as prediction [6]. Many machine learning techniques have been employed in forecasting the potential spread of COVID-19 [5], [11], [13], [14]; however, few studies have reported on the severity prediction of COVID-19 patients [15], [16].

This study discusses the initial prediction of the severity of the COVID-19 stage by clinical information in EHR using machine learning techniques. Early prediction of severity of the COVID-19 stage allows the healthcare organization to develop an effective disease management approach, which may help prevent the progression of the disease. It also helps overcome the limited number of staff in the hospitals by enabling frontliners to help the screening process. By classifying COVID-19 patients into the stage of severity, it helps doctors to prioritized critical patients thus making the screening process more efficient. Further, it may improve the quality of life of the patient. This study also examines the significance of the 21 variables of COVID-19 in the study population. The scope of this study covers the design of the screening system application and the development of machine learning model for the system.

II. METHODOLOGY

A. Project Methodology

We used a hybrid methodology (Fig. 1) which combine waterfall and agile methods. The hybrid methodology model consists of five phases: Requirement, Design, Develop, Test and Evaluation. The requirements and evaluation phases are from the waterfall methodology and the design, test and develop phases are from the agile methodology.

In the requirement phase, we collected the project resources such as dataset, tools and machine learning techniques. We designed the high-level system designs, flowchart and graphical user interfaces in the design phase. In the develop phase, we developed the predictive models using machine learning techniques and we test the predictive models until the best results were achieved. We evaluated the models in the evaluation phase using performance metrics such as accuracy, precision and sensitivity.

B. Model Training Flow

Fig. 2 shows the model training flow that we implemented in the development phase to train and test the machine learning techniques. We used the raw data for data preprocessing, feature selection, train model with training and testing data, and evaluate the model. The data preprocessing involves removing some irrelevant features that does not relevant, removing missing values, removing outliers, data transformation and data balancing. After obtaining the cleaned data, it will be split into two parts which is training data and testing data. The model will learn from the training data while testing data will be used to evaluate the model performance.

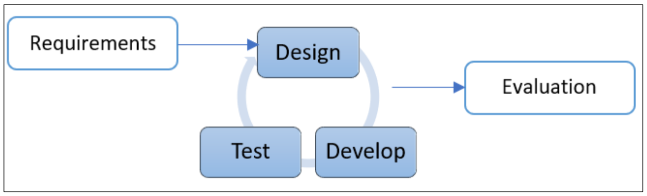


Fig. 1. Hybrid Methodology Model.

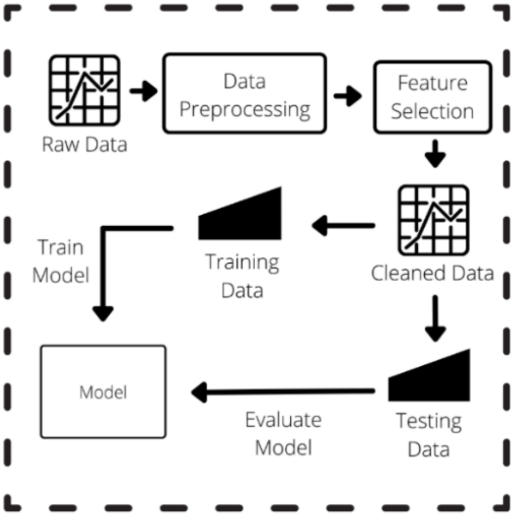


Fig. 2. Model Training Flow.

C. High-Level Screening System Design

The screening system (Fig. 3) consists of two main components: a) system and b) model. The system component requires an input from user to make a screening prediction using the predictive model derived from the model component. The system will response with output (prediction results) and store the results in the database.

D. System Flowchart

The flowchart of the screening system describes the way how the proposed system should work. As shown in Fig. 4, the flow begins with user login. Users need to register their username and password to use the system. When they enter their credentials, the system will verify the user and only allows authenticated users to login into the system. If they fail to verify themselves, the user must input the username and password again. There are two types of users for this system which is the doctor and the nurse. When nurse login, they can enter patient's information and predict the severity of the COVID-19 patient. The system will display the predicted outcome and the score of the severity prediction. The input entered by the nurse will be stored in the database so that doctors can view the data. When the doctor login, the system will show the record of patients and their status. Doctors can keep track of patients' status so that they know which patients are still waiting to be examined and which have been examined.

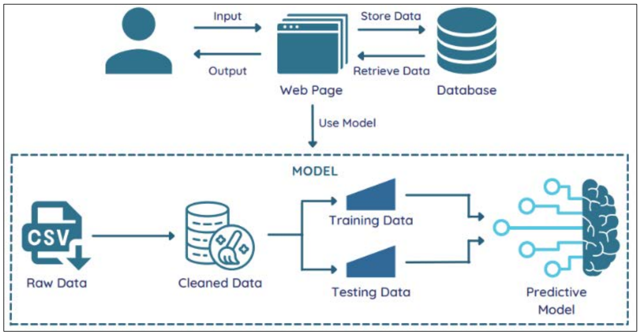


Fig. 3. High-Level Screening System Design.

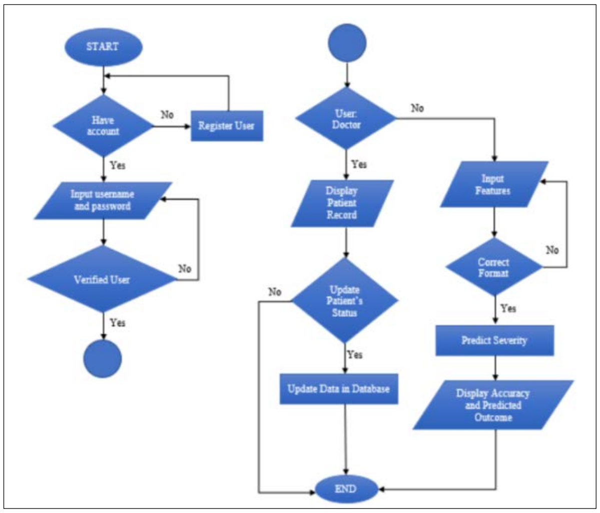


Fig. 4. System Flowchart.

E. Graphical user Interface Design

Fig. 5 shows graphical user interface (GUI) for the screening system that will serve as a way for medical practitioners to interact with the system. The GUI consists of seven web pages with forms for inputs and tables and graph for outputs.

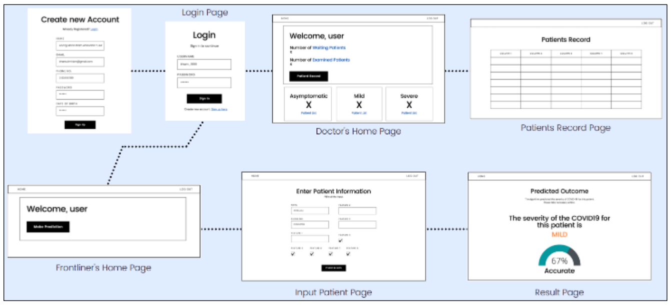


Fig. 5. Proposed Graphical user Interface.

III. Development of Machine Learning Model

A. Dataset

The dataset was obtained from the Springer Nature data repository and consists of 478 patients (https://springernature.figshare.com/articles/dataset/Data_associated_with_the_article_Epidemiological_and_clinical_characteristics_of_imported_cases_with_COVID-19_infection_a_multicentre_study/12159918). The dataset contains the epidemiological and clinical characteristics of COVID-19 cases in China, and consists of 21 variables, including the patient's demographic profile, epidemiological characteristics, clinical data, contact history, case cluster, outcome and type of case. The performance of machine learning models depends upon the quality of the data. Thus, pre-processing was conducted on the dataset to handle missing values, outliers and data generalisation.

After the pre-processing, only eight independent variables that is relevant to COVID-19 symptoms that will be used to predict the dependent variable (type) with three class labels. Table I illustrates the description for each variable.

The original dataset distributions of each class in the target variable are disproportionate which may result in a poor predictive accuracy over the minority class [17] and bias towards the majority class. Thus, to rebalance the data, this study used the synthetic minority oversampling technique (SMOTE) which will increase the minority class to match with majority class. This data balancing technique aims to avoid imbalanced classification in developing predictive models on the dataset. The results before resample and after resample using the SMOTE technique are shown in Fig. 6.

B. Tools

We developed the predictive models using python programming language that provide useful libraries such as Scikit-learn for feature extraction and model training and testing. We designed the GUI forms with text fields and buttons using HTML and CSS. We stored the data in a MySQL database.

C. Pre-processing and Feature Extraction

We analysed and compared the accuracy of three feature selection methods to identify the features or variables that most likely contribute to the severity of illness. The feature selection methods are Pearson Correlation, Random Forest Importance and Recursive Feature Elimination (RFE). Fig. 7 shows the accuracy comparison of feature selection methods applied with the Random Forest model.

TABLE I. VARIABLE DESCRIPTION

Variable	Description
Gender	The gender of the patient.
Age	The patient's age in years.
Fever	Whether the patient have fever.
Cough	Whether the patient have cough.
Fatigue	Whether the patient have fatigue.
Dyspnea	Whether the patient have dyspnea.
Headache	Whether the patient have headache.
H-Temperature	The patient's highest temperature.
Type	The type of cases or the level of severity for the COVID-19 case. (1=Asymptomatic, 2=Mild, 3=Severe)

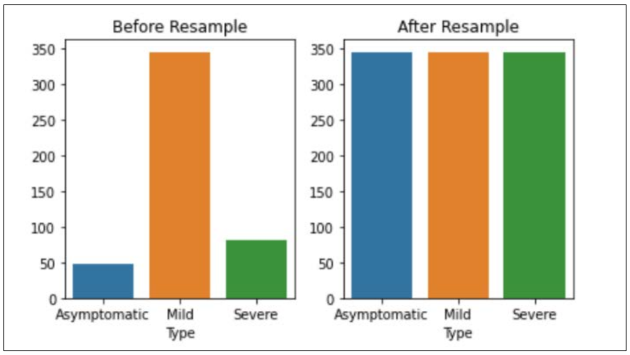


Fig. 6. Original Dataset vs Synthetic Minority Oversampling Technique Dataset.

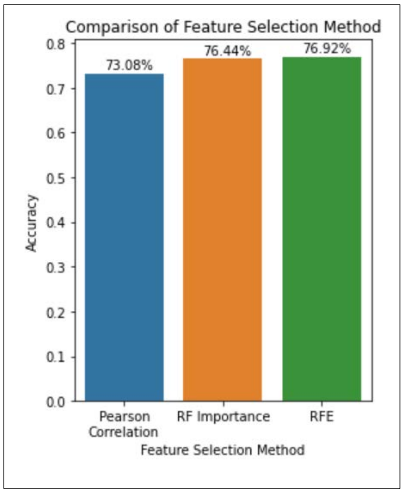


Fig. 7. Comparison of Feature Selection Methods.

The results show that the Random Forest model produces the highest accuracy of 76.92% when using RFE, compared with 76.44% for Random Forest Importance and 73.08% for Pearson Correlation. Thus, RFE was used to determine the importance of the eight features or variables according to the contribution in COVID-19 illness. We used RFE with cross-validation to remove features or variables iteratively and find the most relevant features or variables that contribute to the severity of illness and have strong correlations between the selected features or variables. Fig. 8 shows the accuracy of the model obtained with cross-validation using different numbers of features.

Based on the line chart in Fig. 8, the accuracy of the model is at the highest (78.0%) when the number of features selected is seven. Thus, we used those seven features in this study, which are Gender, Age, Fever, Cough, Fatigue, Dyspnoea and H-Temperature.

D. Model Training

We applied three machine learning techniques: Random Forest (RF), Naïve Bayes (NB) and Gradient Boosting (GB). The brief methodology, working procedure and differences are provided below.

1) Random forest: RF is an ensemble method that combines multiple decision trees (DTs) together in a single forest. It is a powerful supervised machine learning algorithm that is capable of performing both classification and regression tasks with high accuracy [18]–[20]. Various studies in the medical area have used RF algorithms to, for example, diagnose diabetes mellitus [21], [22], identify cervical cancer [23]–[25], or predict the risk of severity for COVID-19 patients at hospital admission [26].

In predicting COVID-19 severity, the RF technique will choose random samples from a given dataset and build an individual decision tree for each sample. One single RF model will have many DTs. Each DT will generate a prediction based on RF technique. As can be seen in Fig. 9, each prediction result contains a vote and the majority votes will be chosen as the final result.

2) Naïve bayes: The NB algorithm works on the Bayesian theorem whereby the probability of a class depends on the probabilities of its variables [27]. The NB classifier is used to maximise the probability of the target class given the features. Previous studies reported that NB can be used to solve the classification problems with multiple classes [28], [29]. In this study, NB can be utilised in the type of COVID-19 severity. Fig. 10 shows the equation used in NB.

3) Gradient boosting: GB is a method to develop classification models by optimising known techniques, such as DT, by adding new learners in a gradual sequential manner [30], [31]. This algorithm is also helpful as a prediction model as it has been used in past research [32], [33] to predict COVID-19. Individually, DT might give a weak prediction ability. However, when combined in an ensemble method, it can improve the accuracy by sequentially upgrading its performance. Fig. 11 shows how boosting in an ensemble method works.

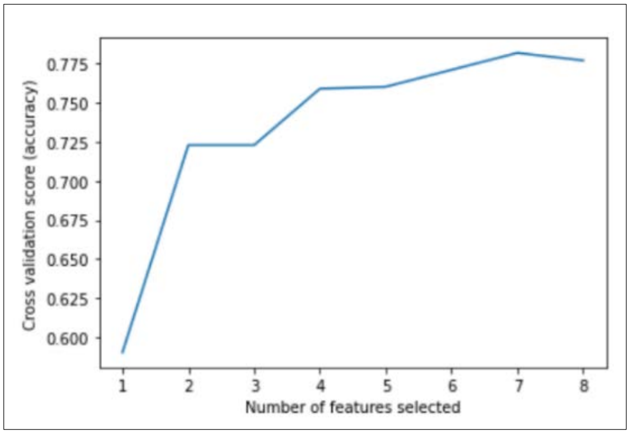


Fig. 8. Accuracy of Model by the Number of Features Selected.

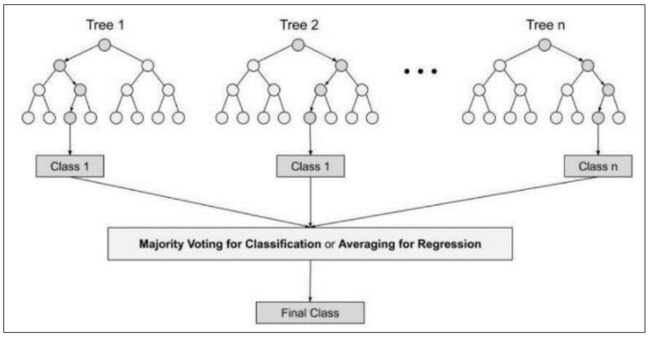

Tree 1
Tree 2
Tree n
Class 1
Class 1
Class n
Majority Voting for Classification or Averaging for Regression
Final Class

Fig. 9. RF Tree.

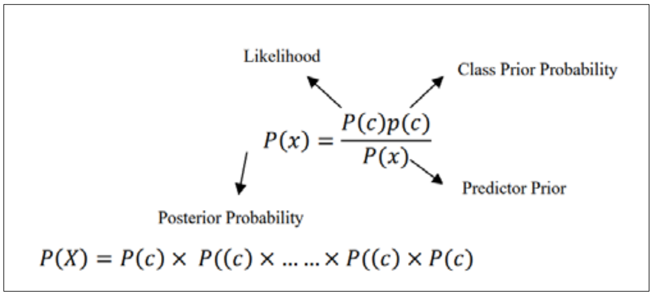

Likelihood
Class Prior Probability
$P(x) = \frac{P(c)p(c)}{P(x)}$
Predictor Prior
Posterior Probability
$P(X) = P(c) \times P((c) \times \dots \dots \times P((c) \times P(c)$

Fig. 10. Equation in NB Classifier [27].

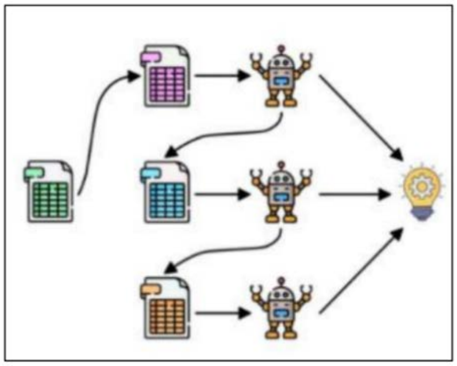


Fig. 11. Boosting Method.

As shown in Fig. 11, although the GB classifier is part of an ensemble method, this method is different from RF in that it does not take the majority vote from each tree. In contrast, it generates the results at a single tree and improves them sequentially until it obtains a good result. The RF, NB, and GB models were run using python software with the abovementioned independent variables or features (Gender, Age, Fever, Cough, Fatigue, Dyspnoea and H-Temperature) and dependent variable (Type), with three class labels.

E. Evaluation Method

We evaluated the performance of machine learning techniques according to evaluation metrics: accuracy, precision, F1-score, recall, sensitivity and specificity. These evaluation metrics used four basic attributes based on confusion matrix which are: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). We calculated the accuracy values using Eq. (1).

$$\text{Accuracy} = TP + TN/(TP + FP + TN + FN) \tag{1}$$

We used precision, recall, F1-score, sensitivity and specificity to support the accuracy values. The method for these metrics is described in Eq. (2), Eq. (3) , Eq. (4) , Eq. (5) and Eq. (6), respectively.

$$\text{Precision} = TP/(TP + FP) \tag{2}$$

$$\text{Recall} = TP/(TP + FN) \tag{3}$$

$$\text{F1-score} = 2 \times (\text{Recall} \times \text{Precision})/(\text{Recall} + \text{Precision}) \tag{4}$$

$$\text{Sensitivity} = TP/(TP + FN) \tag{5}$$

$$\text{Specificity} = TN/(TN + FP) \tag{6}$$

IV. Results

Our main objective is to compare the three machine learning models based on the accuracy performance in this section. In order to analyze the differences, we compare the performance accuracy using the five-fold cross-validation with stratification as a testing method to derive the best predictive model for optimal results. We measure the performance using various metrics including classification accuracy, precision, F1-score, recall, specificity and sensitivity to ensure the predictive model was fit to produce accurate results. Table II shows the performance of the three machine learning models that have been investigated in this study.

Overall, results show that all the machine learning models can be used for predicting the severity of COVID-19 patients. However, the RF model produced the highest accuracy value of 78.4. This may be because of advantages of the RF model, such as building each tree independently and averaging the votes. The averaging method may be advantageous when dealing with multiclassification that aggregates individual predictions into a collective prediction.

The RF and GB models recorded the same values for precision, specificity and sensitivity, which are 91.0, 94.7 and 98.6, respectively. The NB model obtained perfect recall and sensitivity values of 100.

TABLE II. RESULTS OF RF, NB AND GB FOR COVID-19 SEVERITY PREDICTION

Performance measures	**Machine learning model**		
	Random Forest	**Gradient Boosting**	**Naïve Bayes**
Accuracy	**78.4**	**77.5**	**76.1**
Precision	91.0	91.0	89.0
F1-score	95.0	95.0	94.0
Recall	99.0	99.0	100
Specificity	94.7	94.7	93.3
Sensitivity	98.6	98.6	100

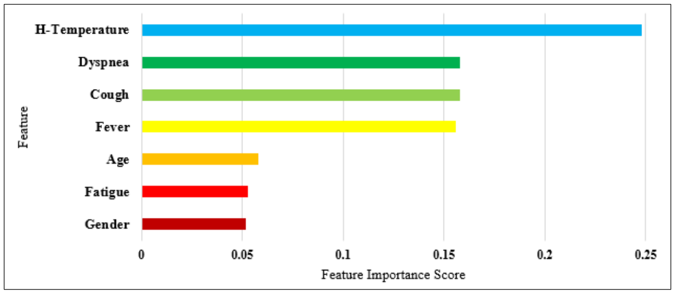


Fig. 12. Visualising Significant Features for COVID-19 Severity Stage Prediction Produced by RF Model.

We conducted another experiment using the RF model to determine the significant features or symptoms in predicting the severity stage of COVID-19 patients. The result is illustrated in Fig. 12, in which the most significant feature is H-Temperature.

V. Discussion

This study set out to report on the performance of machine learning techniques in predicting the severity of COVID-19 patients by using a publicly available dataset from the Springer Nature repository that contains clinical information of COVID-19 patients. Machine learning techniques provide an additional effective way of early screening of patients and do not replace the clinical evaluation.

In our study, results showed that the three machine learning techniques had similar average predictive accuracy in classifying severity of COVID-19 patients (accuracy >0.75). This is consistent with prior findings using an RF model in clinical data [15] and other prior findings using RF, GB, and ensemble learning algorithms in primary care [34]. The RF model outperformed the other machine learning models with the highest accuracy to predict the severity stage of COVID-19.

Despite the similar accuracy, this study found minimum variations for other performance values between the machine learning models. The specificity values were regularly high for RF and GB models, which indicate that the proportion of patients without severe COVID-19 symptoms was correctly classified. The sensitivity and precision values were also consistently high for RF and GB models, indicating that the models identified most of the COVID-19 patients with mild, severe and critical disease severity stages.

The findings of this study have important implications for developing a COVID-19 severity screening system to assist doctors to manage COVID-19 patients before they are examined. Since the outbreak of COVID-19, hospitals have been struggling to keep up with the number of patients because of limited staff. Patients need to be screened so that doctors can prioritise the severe patients as they are at high risk of death. Therefore, more efficient ways to screen the patients are needed because manual screening can be time consuming. By using a COVID-19 severity screening system powered by machine learning models, hospitals can deploy frontliners who are not medical experts to help screen the patients before they undergo actual examination by the doctors. Thus, the time to screen the patients can be shortened.

VI. Conclusion

Machine learning techniques help to reduce the effort and time for medical practitioners to conduct early predictions for healthcare management purposes. As the number of deaths due to COVID-19 increases, a COVID-19 severity screening system is essential to reduce the progression of the disease. In this study, a comparison of three machine learning techniques, RF, NB, and GB were performed, and RF was found to achieve the best performance score. The results of our study show the potential of applying machine learning techniques in the early predictions of a COVID-19 severity screening. Using RF algorithm, body temperature (H temperature) has been found to be important criteria for diagnosing COVID-19 severity level. This may call further investigation to explore temperature data visualization for temperature monitoring of COVID-19 patients. Other than that, a hybrid of machine learning techniques with optimisation algorithms with more data will be examined to further improve accuracy.

Acknowledgment

The authors would like to thank the Ministry of Education Malaysia and Universiti Teknologi MARA (UiTM) for their financial support to this project under UiTM Internal REI Grant 600-IRMI/REI 5/3 (001/2019). We would also like to thank the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Selangor, Malaysia for all the supports.

References

[1] M. E. Hossain, A. Khan, M. A. Moni, and S. Uddin, “Use of Electronic Health Data for Disease Prediction: A Comprehensive Literature Review,” IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 18, no. 2, pp. 745–758, Mar. 2021, doi: 10.1109/TCBB.2019.2937862.

[2] A. J. Steele, S. C. Denaxas, A. D. Shah, H. Hemingway, and N. M. Luscombe, “Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease,” PLoS One, vol. 13, no. 8, 2018, doi: 10.1371/journal.pone.0202344.

[3] S. A.G. et al., “Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma,” American Journal of Gastroenterology, vol. 108, no. 11. 2013.

[4] A. Guo, N. R. Mazumder, D. P. Ladner, and R. E. Foraker, “Predicting mortality among patients with liver cirrhosis in electronic health records with machine learning,” PLoS One, vol. 16, no. 8 August, 2021, doi: 10.1371/journal.pone.0256428.

[5] A. S. Adly, A. S. Adly, and M. S. Adly, “Approaches Based on artificial intelligence and the internet of intelligent things to prevent the spread of COVID-19: Scoping review,” Journal of Medical Internet Research, vol. 22, no. 8. 2020, doi: 10.2196/19104.

[6] K. C. Santosh, “AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data,” J. Med. Syst., 2020, doi: 10.1007/s10916-020-01562-1.

[7] L. S. Khoo, A. H. Hasmi, M. A. Ibrahim, and M. S. Mahmood, “Management of the dead during COVID-19 outbreak in Malaysia,” Forensic Sci. Med. Pathol., 2020, doi: 10.1007/s12024-020-00269-6.

[8] K. Mizumoto and G. Chowell, “Transmission potential of the novel coronavirus (COVID-19) onboard the diamond Princess Cruises Ship, 2020,” Infect. Dis. Model., 2020, doi: 10.1016/j.idm.2020.02.003.

[9] H. Estiri, Z. H. Strasser, J. G. Klann, P. Naseri, K. B. Wagholikar, and S. N. Murphy, “Predicting COVID-19 mortality with electronic medical records,” npj Digit. Med., vol. 4, no. 1, 2021, doi: 10.1038/s41746-021-00383-x.

[10] A. Alotaibi, M. Shiblee, and A. Alshahrani, “Prediction of severity of covid-19-infected patients using machine learning techniques,” Computers, vol. 10, no. 3, Mar. 2021, doi: 10.3390/computers10030031.

[11] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, “Artificial Intelligence (AI) applications for COVID-19 pandemic,” Diabetes Metab. Syndr. Clin. Res. Rev., 2020, doi: 10.1016/j.dsx.2020.04.012.

[12] J. R. Dietz et al., “Recommendations for prioritization, treatment, and triage of breast cancer patients during the COVID-19 pandemic. the COVID-19 pandemic breast cancer consortium,” Breast Cancer Research and Treatment, vol. 181, no. 3. 2020, doi: 10.1007/s10549-020-05644-z.

[13] H. Estiri, Z. H. Strasser, J. G. Klann, P. Naseri, K. B. Wagholikar, and S. N. Murphy, “Predicting COVID-19 mortality with electronic medical records,” npj Digit. Med., vol. 4, no. 1, Dec. 2021, doi: 10.1038/s41746-021-00383-x.

[14] A. Kumar, P. K. Gupta, and A. Srivastava, “A review of modern technologies for tackling COVID-19 pandemic,” Diabetes Metab. Syndr. Clin. Res. Rev., 2020, doi: 10.1016/j.dsx.2020.05.008.

[15] A. Alotaibi, M. Shiblee, and A. Alshahrani, “Prediction of severity of covid-19-infected patients using machine learning techniques,” Computers, vol. 10, no. 3, 2021, doi: 10.3390/computers10030031.

[16] W. A. Abbasi et al., “COVIDC: An expert system to diagnose COVID-19 and predict its severity using chest CT scans: Application in radiology,” Informatics Med. Unlocked, vol. 23, Jan. 2021, doi: 10.1016/j.imu.2021.100540.

[17] L. Wei, W. Luo, J. Weng, Y. Zhong, X. Zhang, and Z. Yan, “Machine Learning-Based Malicious Application Detection of Android,” IEEE Access, vol. 5, pp. 25591–25601, 2017, doi: 10.1109/ACCESS.2017 .2771470.

[18] R. Srivatsan, P. N. Indi, S. Agrahari, S. Menon, and S. D. Ashok, “Machine learning based prognostic model and mobile application software platform for predicting infection susceptibility of COVID-19 using healthcare data,” Res. Biomed. Eng., vol. 38, no. 1, pp. 59–70, Mar. 2022, doi: 10.1007/s42600-020-00103-6.

[19] N. Amoroso, R. Cilli, T. Maggipinto, A. Monaco, S. Tangaro, and R. Bellotti, “Satellite data and machine learning reveal a significant correlation between NO2 and COVID-19 mortality,” Environ. Res., vol. 204, p. 111970, Mar. 2022, doi: 10.1016/j.envres.2021.111970.

[20] I. Alam, D. Md. Farid, and R. J. F. Rossetti, “The prediction of traffic flow with regression analysis,” in Advances in Intelligent Systems and Computing, 2019, vol. 813, doi: 10.1007/978-981-13-1498-8_58.

[21] X. Wang et al., “Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier,” BMC Med. Inform. Decis. Mak., vol. 21, no. 1, p. 105, Dec. 2021, doi: 10.1186/s12911-021-01471-4.

[22] N. Komal Kumar, D. Vigneswari, M. Vamsi Krishna, and G. V. Phanindra Reddy, “An Optimized Random Forest Classifier for Diabetes Mellitus,” in Advances in Intelligent Systems and Computing, vol. 813, 2019, pp. 765–773.

[23] K. P. Win, Y. Kitjaidure, K. Hamamoto, and T. Myo Aung, “Computer-Assisted Screening for Cervical Cancer Using Digital Image Processing of Pap Smear Images,” Appl. Sci., vol. 10, no. 5, p. 1800, Mar. 2020, doi: 10.3390/app10051800.

[24] S. F. Abdoh, M. Abo Rizka, and F. A. Maghraby, “Cervical Cancer Diagnosis Using Random Forest Classifier With SMOTE and Feature Reduction Techniques,” IEEE Access, vol. 6, pp. 59475–59485, 2018, doi: 10.1109/ACCESS.2018.2874063.

[25] D. N. Diniz et al., “A Hierarchical Feature-Based Methodology to Perform Cervical Cancer Classification,” Appl. Sci., vol. 11, no. 9, p. 4091, Apr. 2021, doi: 10.3390/app11094091.

[26] G. Wu et al., “Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: An international multicentre study,” Eur. Respir. J., vol. 56, no. 2, 2020, doi: 10.1183/13993003.01104-2020.

[27] D. Berrar, “Bayes’ theorem and naive bayes classifier,” in Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, vol. 1–3, 2018.

[28] L. Yahaya, N. David Oye, and E. Joshua Garba, “A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques,” Am. J. Artif. Intell., vol. 4, no. 1, 2020, doi: 10.11648/j.ajai.20200401.12.

[29] W. N. L. W. H. Ibeni, M. Z. M. Salikon, A. Mustapha, S. A. Daud, and M. N. M. Salleh, “Comparative analysis on bayesian classification for breast cancer problem,” Bull. Electr. Eng. Informatics, vol. 8, no. 4, 2019, doi: 10.11591/eei.v8i4.1628.

[30] N. Chakrabarty, T. Kundu, S. Dandapat, A. Sarkar, and D. K. Kole, “Flight arrival delay prediction using gradient boosting classifier,” in Advances in Intelligent Systems and Computing, 2019, vol. 813, doi: 10.1007/978-981-13-1498-8_57.

[31] A. Pina et al., “Virtual genetic diagnosis for familial hypercholesterolemia powered by machine learning,” Eur. J. Prev. Cardiol., 2020, doi: 10.1177/2047487319898951.

[32] M. Kukar et al., “COVID-19 diagnosis by routine blood tests using machine learning,” Sci. Rep., vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-90265-9.

[33] Y. Zoabi, S. Deri-Rozov, and N. Shomron, “Machine learning-based prediction of COVID-19 diagnosis based on symptoms,” npj Digit. Med., vol. 4, no. 1, 2021, doi: 10.1038/s41746-020-00372-6.

[34] R. F. Albuquerque Paiva de Oliveira, C. J. A. Bastos Filho, A. C. A. M. V. F. de Medeiros, P. J. Buarque Lins dos Santos, and D. Lopes Freire, “Machine learning applied in SARS-CoV-2 COVID 19 screening using clinical analysis parameters,” IEEE Lat. Am. Trans., vol. 19, no. 6, pp. 978–985, Jun. 2021, doi: 10.1109/TLA.2021.9451243.

Electronic Personal Health Record Assessment Methodology: A Review

Dirayana Kamarudin, Nurhizam Safie, Hasimi Sallehudin
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
Selangor, Malaysia

***Abstract*—ePHR (Electronic Personal Health Record) is not a new concept in the era of electronic health information. The advantages of ePHR in improving health outcomes through patient empowerment have been recognized globally and almost all countries that implement electronic health records (EHR) have created ePHR. This study identifies the components of the ePHR implementation study methodology that has been conducted throughout the country. The types of ePHR studies selected were adoption studies, acceptance studies, readiness studies, and evaluation studies. This study's systematic literature review process is identification, screening, eligibility, data abstraction, and analysis. A total of 16 final journals were analyzed from 173 journals identified from 5 databases (Science Direct, WoS, Scopus, JMIR, and PubMed) regardless of the year of publication until April 1st, 2021. Among the findings based on the four objectives of the study, there are two findings that are considered important and interesting by the author; first, the existence of 22 additional variables to the evaluation model by almost all studies in this study which shows a clear need to improve the evaluation model which is the TAM Model. Second, although the proposal of conducting a scientific study to evaluate the perspective of ePHR stakeholders before ePHR is developed only appeared once, based on this study and the knowledge of the authors, it is a starting point for the successful implementation of ePHR. These two findings contribute to the recommendations for the best design of the ePHR implementation study described in this paper.**

Keywords—Personal health record; ePHR; ePHR evaluation variable

I. Introduction

A. PHR (Personal Health Record)

PHR is part of health information. Health Information can provide five supporting powers to individuals, namely prevention, treatment, protection, health service and resource planning [1]. A physician knows the more knowledge about a patient, the higher the patient's chances of receiving the best possible health care. Patients also need adequate health resources to recover more quickly while receiving fewer complications from the optimal treatment [2]. The increasing population equipped with new and complex health treatments has increased the demand for better and more efficient health services around the world. Complicated health problems have created more complex health care processes. The need for complete health information for better treatment is also increasing. Now, it can be seen that the increase in Electronic Health Record (EHR) initiatives is one of the methods of winning this issue [3], [4]. PHR emerges from an EHR and is defined as a health record related to a patient-administered treatment [5]. PHR is a general data set of individual lifelong health information that the public understands that can be securely [6] accessed at any time for the purpose of treatment (illness) and wellness. PHR is owned and managed entirely by the individual or a party appointed by that individual. PHR can exist in the form of electronic health records that comply with recognized interoperability standards and can be sourced from a various sources while accessed, shared, and operated by individuals [7]. EHRs are generated throughout the patient's engagement with healthcare-related parties. This information consists of patient demographic data, patient progress records, patient problems, patient medications, patient health reading level information, patient medical history, patient immunizations, laboratory test result data and patient x-ray report. EHRs automate the workflow of physicians. EHR has some limitations because its records are based entirely on data reported by healthcare providers [8], [9] resulting in the existence of a trend that allows patients to gain access to their health data and makes them the owners of such data called PHR [10]. Based on the literacy study [10], PHR and EHR have different purposes, namely PHR is for the personal domain and EHR is for the organizational domain. The PHR is used for self-health management and monitoring that can be collaborated with the patient's digital device. At the same time PHR can integrate with EHR, EMR and other systems such as health insurance systems. PHR and EHR also can be integrated to exchange relevant patient health information [10]. Electronic Medical Record (EMR) is an application that consists of a repository of clinical data, clinical decision support, medical terminology, order requests by staff and medical practitioners and clinical documentation applications [7]. This environment supports patients' EMR in inpatient and outpatient environments and is used by healthcare practitioners to document, monitor, and manage health delivery services. Individual health information can be created, collected, managed and consulted by physicians and staff of healthcare organizations [7].

B. PHR Implementation

Based on this studies, it was found that PHR has existed since 1973 [11] and ePHR has existed since 2001 [12]. PHR is classified based the on health service provider, user type and system channel [5]. PHR user profiles are patients/individuals, health professionals and authorized third parties (patient/ individual families and government parties) who perform the process of consultation, monitoring and maintenance of individual health [10]. PHR exists in the form of paper, and computer systems integrated between several health facilities and hybrids (computer systems that can be accessed anywhere) [10]. There are 20 types of data (Allergies, Demographic, Documents, Evolution, Family History, General, Genetic, Home Monitor, Immunizations, Insurance, Laboratory Results, Major Illness, Medications, Prescriptions, Prevention, Providers, Scheduling, Social History, Summaries and Vital Signs) that exist in the PHR and half of them (Documents, Evolution, Immunizations, Insurance, Laboratory Results, Medications, Prescriptions, Scheduling, Summaries and Vital Signs) exist in the EHR [10].

C. Implementation of ePHR in Malaysia

Until April 1st, 2021, its shows that Malaysia has not yet conducted any ePHR acceptance study. Referring to the Information Technology Strategic Plan of the Ministry of Health Malaysia 2016-2020 [13], ePHR is not listed as one of the initiatives to be implemented. However, there is a journal [14] that shows the production of variable models (UTAUT2 and PMT) to study of ePHR acceptance in Malaysia. Still, the authors found no references to show the results of studies using such models. Next, there is a study [15] prepared for the Ministry of Health Malaysia (MoHM) to develop a Lifetime Health Record (LHR) system [13]. The study yielded the structure of the LHR dataset. This dataset is prepared for the MoHM based on the clinical consultation process and the use of patient demographic records in the clinic. This LHR dataset is provided according to the needs of healthcare professionals. The proposed LHR has three components namely Patient Master Information, Health Condition Summary and Episode Summary. This research concludes that ePHR needs to be developed by taking into account the structure and components of LHR records according to the needs and perspectives of Malaysians. Also, there was last finding a study [16] about LHR in hospital and clinic of Ministry of Defense Malaysia.

D. Need for the Review

According to Alsahafi [5] the implementation of the health record computer system is moving from a service provider control information system to a patient/individual control information system, namely, ePHR (electronic PHR). He also noted that several countries such as the United States, England and Australia have chosen to implement globally integrated ePHR to improve the quality of health service delivery. The results of his study concluded that, although the advantages of ePHR are agreed almost worldwide, the acceptance rate of the implementation of ePHR is still at a low level. The authors have agreed with HS Park's suggestion that research is needed before ePHR is developed to identify the real needs and concerns of those interested in ePHR [17] and thus be one of the ways to solve the problem of low ePHR acceptance rate. Based on the reading of the journal until April 1st, 2021, the authors have agreed that a systematic review of the literature on the methodology of existing ePHR implementation studies (readiness, adoption, adaptation and evaluation) is needed to determine the best study design for ePHR studies.

II. Objective of Study

The objectives of the study to be archived sequentially are as follows:

1) To identify the scope of the ePHR implementation study: The authors define the meaning of the scope of ePHR implementation as a study of two research boundaries, namely studies before or after the implementation of ePHR and the categories of ePHR study respondents such as patients, public, health professionals and others.

2) To identify ePHR implementation research methodology: The methodological component chosen in this study is the method of collecting and analyzing research data which is qualitative, quantitative or mixed methods. Next, the authors want to see, the method is categorized according to the scope of the study that has been identified when achieving the first objective.

3) To identify ePHR implementation acceptance models and the results: The authors want to identify the models that exist in ePHR implementation studies including the variables that exist in those studies. Next, the authors wanted to obtain all results from all selected studies.

4) To identify ePHR implementation study recommendations: The authors want to get specific improvement suggestions for the implementation of ePHR that exist and are not bound to the scope of the study that has been identified in the achievement of the first objective.

III. Methodology

Referring to systematic literature review study by Mohamed Shaffril [18], the systematic literature review process used for this study follow the same process which are identification, screening, eligibility, data abstraction and analysis.

A. Identification

To identify the journal topic for this study, the search was made based on the review searching topic map as in Fig. 1.

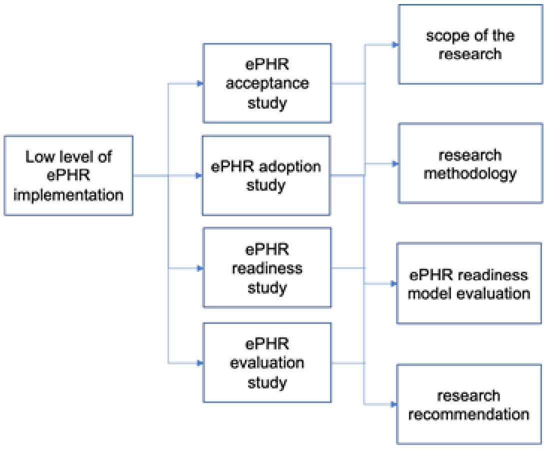


Fig. 1. Review Searching Topic.

Journal searches based on the topics in Fig. 1 were made in the databases of Science Direct (SD), World of Science (WoS), Scopus, Journal of Medical Internet Research (JMIR) and PubMed. As of April 1, 2021, the number of journals obtained based on the selected keywords and databases are as shown in Table I. The total number of a document identified was 173. The keywords used throughout this study are as follows:

1) Keyword A = "Personal Health Record" AND "Acceptance"

2) Keyword B = "Personal Health Record" AND "Adoption"

3) Keyword C = "Personal Health Record" AND "Readiness"

4) Keyword D = "Personal Health Record" AND "Evaluation"

TABLE I. NUMBER OF JOURNAL BY KEYWORD AND DATABASE

	SD	**WoS**	**Scopus**	**JMIR**	**PubMed**	Total
Keyword A	0	12	11	1	3	27
Keyword B	3	30	41	3	9	86
Keyword C	0	1	1	0	0	2
Keyword D	2	23	21	1	11	58
Total	5	66	74	5	23	173

B. Screening

The first screening is a review of the title and author of the document. Duplicate documents will be removed from the list. Next, a second screening was made by selecting published documents as journals and published in English. The journal is also a journal that is in the final stages. The screening results are as in Table II. Since the document identification results show that studies on the implementation of PHR are still lacking, this study does not specify the year of publication of the journal. The final number of journals after the second screening was 53.

TABLE II. NUMBER OF JOURNAL BY KEYWORD AFTER SCREENING

Keyword	**Journal**	**Total**
Keyword A	[19], [17], [20], [21], [14], [22], [23], [24],	**8**
Keyword B	[25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50]	**26**
Keyword C	[51]	**1**
Keyword D	[52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [59], [62], [63], [64], [65], [66], [67], [68]	**18**
Total Journal		**N=53**

C. Eligibility

At this stage, the journal abstract is read carefully to identify the scope, methodology, evaluation model, results and recommendations of the study. If such information is not available in the abstract, a reading will be performed on the main document of the journal. Journals containing incomplete information are removed for further processing. Finally, a total of 16 articles as in Appendix I were selected for analysis. The percentage of the selected journal that provides sufficient information for the study is described in Table III. It shows only 9% of the journals found could be analyzed.

TABLE III. NUMBER AND PERCENTAGE OF SELECTED JOURNAL FOR THE STUDY

Identification			**Screening**			**Eligibility**
Keyword A						
27	Y	8	8	Y	6	6
	N	19		N	2	
Keyword B						
86	Y	26	26	Y	8	8
	N	60		N	18	
Keyword C						
2	Y	1	1	Y	0	0
	N	1		N	1	
Keyword D						
58	Y	18	18	Y	2	2
	N	40		N	16	
173	Y	53 (31%)	53	Y	16 (30%)	16 (9% of journal found)
	N	120 (69%)		N	37 (70%)	

Notes:
Y = Selected journal after certain review process
N = Rejected journal after certain review process

D. Data Abstraction and Analysis

The methods of analysis used in this study are as follows:

1) Data extraction: Data related to the five major themes were extracted from the journal's primary documents. The themes are the scope, methodology, evaluation model, results and recommendations of the study.

2) Data grouping: Data for the scope, methodology and evaluation model were consolidated and grouped according to sub-themes as in Appendix I. The original data for the results and recommendations were summarized and grouped according to the themes of the data as in Appendix II.

IV. FINDINGS

The findings of the study were compiled based on the objectives of the study.

A. Study about the Scope of the ePHR Implementation

Based on the analysis, it was found that 9 (56%) studies were implemented before ePHR was implemented. 8 (50%) studies were conducted before ePHR was implemented. 10 (62%) studies made patients as respondents, 5 (31%) studies used health professionals, 3 (19%) studies used the general public and 1 (6%) studies used health organization management staff. Only 1 of the 16 studies incorporated 3 groups of respondents: patients, health professionals and management staff of health organizations and the study was conducted before the implementation of ePHR. All studies (3 studies) involving the general public were evaluated before ePHR was implemented. Meanwhile, 6 out of 10 studies involving patients were evaluated after the implementation of ePHR.

B. Study about the Research Methodology for ePHR Implementation

It was found that 11 (69%) studies were conducted with a quantitative method, 3 (19%) studies with mixed-method and 2 (12%) studies with qualitative method. For the studies conducted before implementation of ePHR, only 7 (78%) studies were conducted quantitatively and the remaining 2 (22%) studies were conducted qualitatively. Furthermore, for the studies conducted after the implementation of ePHR, 4 (50%) studies were conducted quantitatively, 3 (38%) studies with mixed methods and 1 (12%) studies were conducted qualitatively. It is found that all studies that make the public as respondents use a similar method that is quantitative methods. Meanwhile, 6 (60%) studies on patients were conducted quantitatively, the remaining 3 (30%) studies were conducted qualitatively and 1 (10%) studies were mixed method.

C. Study about the Acceptance Models and Results for ePHR

It was found that eight (50%) studies adapted the TAM (Technology Acceptance Model) model as the evaluation model, two (12%) studies used the UTAUT (Unified Theory and Use of Technology) model and 6 (38%) studies used the SDT (Self-Determination Theory) model. Of the 16 studies, only 1 study used the TAM model completely without additional variables. The remaining 15 (94%) studies selected additional variables based on researcher's knowledge named SDT. Of the 15 studies, 7 (47%) studies were combined with the TAM model, 2 (13%) studies were combined with the UTAUT model and the remaining 6 (40%) studies used the SDT model entirely. TAM combined variable model developed by Fred Davis in 1989 based on TRA (Theory of Reasoned Action) by Ajzen and Fishbein [24]. TAM contains two main factors of acceptance of a technology: Perceived Usefull (PU) and Perceived Ease-of-Use (PEOU). Venkatesh developed UTAUT in 2003 due to the integration of 8 models [69]. UTAUT has three factors that influence Behavioral Intention (BI) Factors: Performance Expectancy, Effort Expectancy, and Social Influence. Two factors that influence Use Behavior Factors are BI and Facilitating Conditions. There are four types of moderators, namely, Gender, Age, Experience and Voluntariness of Use. The SDT variables that has identified by this study are as shown in Table IV.

Referring to Appendix II, the results of the study are summarized into thhree themes, namely, all IVs have an effect on DV, some IVs do not have an effect on DV and the number of themes produced. It was found that almost all studies gave the same results, that is, all the variables (IVs and themes) studied have been scientifically proven to affect the implementation of ePHR.

D. Study about Recommendations for ePHR Implementation

Appendix II contains recommendations from 16 studies summarized into 8 themes and taken the total frequency of the recommendations were raised. Table V shows the list of suggestions arranged according to most frequently mentioned.

A summary of all the findings categorized according to the first three objectives of the study is as in Appendix III.

TABLE IV. SDT VARIABLE BY QUALITATIVE STUDY

SDT Variable (Quantitative Study)	Journal
1. Confidentiality 2. Privacy	[27]
3. e-Health Literacy	[19]
4. Physician Autonomy Support 5. Autonomous Causality Orientation 6. Basic Needs Satisfaction	[32]
7. Health-care technology self-efficacy	[31]
8. Perceived data privacy 9. Security protection 10. Perceived health-promoting role model	[22]
11. Subjective Norm, 12. Security & Privacy 13. Computer self-efficacy	[43]
14. Task Technology Fit 15. Patient Activation Measure	[33]
16. Physician-patient Relationship	[23]
17. Technology Barriers	[24]
18. Perceived Risk 19. Facilitating Conditions	[17]
20. Compatibility 21. Communicative	[70]
22. Impact on current workflow	[38]

TABLE V. LIST OF RECOMMENDATION

Percentage of frequency	Recommendation
62%	1) Increase ePHR awareness or education among stakeholder
50%	2) The health proffesional should play a central role to improve utility and consequently the adoption of the ePHR
31%	3) Design ePHR application acording to stakeholder need and concern
19%	4) Policymakers, and health-care providers must pay additional attention to increasing individuals conviction and confidence in using the ePHR
6%	5) Conduct sufficient study to identify stakeholders' perspective and need before ePHR development/ implementation.
6%	6) Healthcare policy-makers, physicians, and developers must consider actions to improve the usability of ePHR in the future
6%	7) Some legal and ethical issues also need to be considered for ePHR adoption
6%	8) Improve the quality of the physician-patient relationship

V. Discussion of Result

Based on the findings of the study, the matters to be discussed of result are divided into three, namely the improvement of the evaluation model, the selection of key recommendations and the best design for ePHR implementation study.

A. Improvement of ePHR Evaluation Model with SDT

Referring to Appendix I, 69% (11 out of 16) of the studies were carried out quantitatively using the popular model that is TAM or UTAUT and 91% (10 out of 11) of the studies created an additional variable (SDT) that was combined with the existing model. These additional variables are scientifically proven to impact ePHR implementation. This shows that there is a need for improvement of either TAM or UTAUT models to evaluate health digital initiatives and researchers cannot wait for these improvements and are forced to develop their own SDT. The scientific selection of SDT was not addressed clearly in most of the research paper. It was found that the additional variables used by the researchers were based on the knowledge of the researchers [27], [31].

Based on the findings of this study, it can be concluded that there is a significant scenario of adding variables to the TAM and UTAUT models. Therefore, researchers who want to study the implementation of ePHR and want to identify the appropriate evaluation model, the authors suggest that they refer to the list of SDT variables in Table IV to get broad ideas for determining additional variables that want to be combined with the TAM or UTAUT model. All SDT variables listed have been scientifically proven by researchers and almost all of these variables have been evaluated from the patient's point of view. Researchers who want to improve the TAM or UTAUT Model or want to create a new model, researchers can also refer to Table IV to develop the most suitable model to evaluate the implementation of digital health initiatives so that the initiative is finally accepted and used fully.

B. Selection of Key Recommendation of ePHR Implementation

Although the awareness program is the most frequently mentioned recommendation, but the authors strongly agreed that the fifth recommendation in Table VI is the main and best recommendation that should be considered by all parties involved in the implementation of ePHR. Directly, the fifth recommendation is the first step to the success of the third recommendation expressed by 31% of the study. The fifth recommendation can indirectly help implementation the remaining recommendations.

C. The Best Design for ePHR Implementation Study

From these two findings, 9/16 studies (56%) were conducted before the ePHR was implemented and 6% conducted sufficient studies to identify the perspectives and needs of stakeholders before the development/implementation of the ePHR, the authors argue that although the studies before and after did not show significant differences in the pattern of results and percentage of the population, there are research recommendations before the development/implementation of ePHR that are selected as key recommendations as described in previous sub topic that can be taken seriously by ePHR implementation researchers.

It was found that seven out of nine (78%) studies conducted before the implementation of ePHR, were carried out qualitatively, since the majority of studies carried out before the implementation of ePHR were carried out qualitatively, this study uses the same approach. It is in line with the concept of research design [71] which is exploratory towards the perspective of the general public, patients and health professionals towards the development/implementation of ePHR which is more suitable to be carried out qualitatively.

As explained in sub-topic V(A), the ePHR implementation study is thought to be most suitable to be carried out by using the TAM model which is improved with SDT variables as in Table IV which can be studied on patients, the public and health professionals together or separately.

VI. Limitation and Recommendations

From the point of view of digital initiative evaluation theory, this study only focuses on the evaluation of ePHR implementation. The authors chose to identify the components of the research methodology specific to ePHR before answering the question of whether the components of the ePHR evaluation model are similar to the digital initiative evaluation model in other areas such as finance, transportation and so on. Before comparisons can be made, a review of digital initiative evaluation studies in other fields must be conducted. The next question that may be studied is the model of evaluating health digital initiatives from the point of view of the public, patients, health professionals and health service providers as health digital initiatives have now changed from professional-centric to patient-centric.

VII. Conclusion

It is hoped that the results of this study can help those who want to develop, implement and evaluate ePHR. The authors believe that the new variables listed are realistic and up-to-date variables as well as recommendations that have been identified as relevant to be implemented for the successful implementation of ePHR.

Acknowledgment

We gratefully acknowledge that this work supported by the Ministry of Higher Education of Malaysia under the FRGS/1/2020/TK0/UKM/02/9 and Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia. We would also like to acknowledge the editors and anonymous reviewers for their valuable guidance.

References

[1] R. Michael, Vision and Value in Health Information. Taylor & Francis Group, 2003. doi: 10.1201/9781315375380.

[2] E. Coiera, GUIDE TO HEALTH INFORMATICS, 2nd ed. Hodder Arnold, 2003.

[3] Pradeep Sinha et al., Electronic Health Record. IEEE Press & WILEY, 2013.

[4] N. A. Ahayalimudin, “An Overview of Electronic Medical Record Implementation in Healthcare System : Lesson to Learn”, doi: 10.5829/idosi.wasj.2013.25.02.2537.

[5] A. Y. A. Alsahafi and B. V. Gay, "An overview of electronic personal health records," Heal. Policy Technol., vol. 7, no. 4, pp. 427–432, 2018, doi: https://doi.org/10.1016/j.hlpt.2018.10.004.

[6] A. Ghazvini and Z. Shukur, "Security Challenges and Success Factors of Electronic Healthcare System," Procedia Technol., vol. 11, no. Iceei, pp. 212–219, 2013, doi: 10.1016/j.protcy.2013.12.183.

[7] G. O. Christine Hudak, HIMSS Dictionary of Health Information and Technology Terms, Acronyms, and Organizations. Taylor & Francis Group, 2019.

[8] A. Ismail, A. T. Jamil, A. F. A Rahman, J. M. Abu Bakar, N. Mohd Saad, and H. Saadi, "The implementation of Hospital Information System (HIS) in tertiary hospitals in Malaysia," Malaysian J. Public Heal. Med. 2010, vol. 10, no. 2, pp. 16–24, 2010.

[9] A. Aman, "Clinical information systems in private hospitals," Int. Conf. Adv. Commun. Technol. ICACT, pp. 1032–1035, 2013.

[10] A. Roehrs, C. A. Da Costa, R. Da Rosa Righi, and K. S. F. De Oliveira, "Personal health records: A systematic literature review," J. Med. Internet Res., vol. 19, no. 1, 2017, doi: 10.2196/jmir.5876.

[11] T. Okawa, "A personal health record for young female students," [Josanpu zasshi] Japanese J. midwife, vol. 27, no. 11, pp. 36–40, 1973.

[12] I. C. Denton, "Will patients use electronic personal health records? Responses from a real-life experience.," J. Healthc. Inf. Manag., vol. 15, no. 3, pp. 251–259, 2001.

[13] Ministry of Health Malaysia, pelan strategik teknologi maklumat KKM 2016-2020. 2016.

[14] A. Mamra, A. Samad, G. Pramudya, M. Bader, Y. Hamad, and M. Doheir, "A Proposed Framework to Investigate the User Acceptance of Personal Health Records in Malaysia using UTAUT2 and PMT," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 3, 2017, doi: 10.14569/ijacsa.2017.080353.

[15] M. Khanapi, A. Ghani, R. K. Bali, R. N. G. Naguib, and I. M. Marshall, "The Analysis and Design of a Pervasive Health Record : Perspectives From Malaysia," in Healthcare Delivery in the Information Age, Springer, 2013. doi: 10.1007/978-1-4614-4514-2.

[16] H. Sallehudin, A. F. Fadzil, and R. Baker, "INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION A Conceptual Study of User Adoption for Military Lifetime Health Record Systems," vol. 3, pp. 75–78.

[17] H. S. H. S. Park et al., "Factors Influencing Acceptance of Personal Health Record Apps for Workplace Health Promotion: Cross-Sectional Questionnaire Study.," JMIR mHealth uHealth, vol. 8, no. 6, p. e16723, 2020, doi: 10.2196/16723.

[18] H. A. Mohamed Shaffril, A. A. Samah, S. F. Samsuddin, and Z. Ali, "Mirror-mirror on the wall, what climate change adaptation strategies are practiced by the Asian's fishermen of all?," J. Clean. Prod., vol. 232, pp. 104–117, 2019, doi: 10.1016/j.jclepro.2019.05.262.

[19] Y. A. Alsahafi, V. Gay, and A. A. Khwaji, "Factors affecting the acceptance of integrated electronic personal health records in Saudi Arabia: The impact of e-health literacy," Heal. Inf. Manag. J., 2020, doi: 10.1177/1833358320964899.

[20] M. Cocosila and N. Archer, "Modeling consumer acceptance of electronic personal health records," J. Electron. Commer. Res., vol. 19, no. 2, pp. 119–134, 2018.

[21] J. Razmak and C. Bélanger, "Using the technology acceptance model to predict patient attitude toward personal health records in regional communities," Inf. Technol. People, vol. 31, no. 2, pp. 306–326, 2018, doi: 10.1108/ITP-07-2016-0160.

[22] K. Gartrell, A. M. Trinkoff, C. L. Storr, M. L. Wilson, and A. P. Gurses, "Testing the electronic personal health record acceptance model by nurses for managing their own health: A cross-sectional survey," Appl. Clin. Inform., vol. 6, no. 2, pp. 224–247, 2015, doi: 10.4338/ACI-2014-11-RA-0107.

[23] C.-F. Liu, Y.-C. Tsai, and F.-L. Jang, "Patients' acceptance towards a web-based personal health record system: An empirical study in Taiwan," Int. J. Environ. Res. Public Health, vol. 10, no. 10, pp. 5191–5208, 2013, doi: 10.3390/ijerph10105191.

[24] A. M. Noblin, T. T. H. Wan, and M. Fottler, "Intention to use a personal health record: A theoretical analysis using the technology acceptance model," Int. J. Healthc. Technol. Manag., vol. 14, no. 1–2, pp. 73–89, 2013, doi: 10.1504/IJHTM.2013.055085.

[25] C. C. Yousef et al., "Adoption of a Personal Health Record in the Digital Age: Cross-Sectional Study," J. Med. Internet Res., vol. 22, no. 10, 2020, doi: 10.2196/22913.

[26] A. Abd-Alrazaq, A. A. A. A. Alalwan, B. McMillan, B. M. B. M. Bewick, M. Househ, and A. T. A. T. A. T. Al-Zyadat, "Patients' Adoption of Electronic Personal Health Records in England: Secondary data analysis (Preprint)," J. Med. Internet Res., vol. 22, no. 10, p. e17499, 2020, doi: 10.2196/17499.

[27] M. Abdekhoda, A. Dehnad, and H. Khezri, "The effect of confidentiality and privacy concerns on adoption of personal health record from patient's perspective," Health Technol. (Berl)., vol. 9, no. 4, pp. 463–469, 2019, doi: 10.1007/s12553-018-00287-z.

[28] A. Alanazi and Y. A. Y. Al Anazi, "The Challenges in Personal Health Record Adoption," J. Healthc. Manag., vol. 64, no. 2, pp. 104–109, 2019, doi: 10.1097/JHM-D-17-00191.

[29] B. Sterud, "Practitioner application: The Challenges in Personal Health Record Adoption," J. Healthc. Manag., vol. 64, no. 2, pp. 109–110, 2019, doi: 10.1097/JHM-D-19-00010.

[30] I. Badran, A. Bruynseels, S. Khan, F. Sii, and P. Shah, "Barriers to adoption of a personal health record in an ophthalmic setting: lessons from implementation of a Glaucoma Patient Passport.," Clin. Ophthalmol., vol. 13, pp. 1369–1375, 2019, [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&db=mdc&AN=31440025&site=ehost-live.

[31] B. Dutta, M.-H. M.-H. Peng, and S.-L. S.-L. Sun, "Modeling the adoption of personal health record (PHR) among individual: the effect of health-care technology self-efficacy and gender concern.," Libyan J. Med., vol. 13, no. 1, p. 1500349, 2018, doi: 10.1080/19932820.2018.1500349.

[32] V. Assadi and K. Hassanein, "Consumer adoption of personal health record systems: A self-determination theory perspective," J. Med. Internet Res., vol. 19, no. 7, 2017, doi: 10.2196/jmir.7721.

[33] J. Laugesen and K. Hassanein, "Adoption of personal health records by chronic disease patients: A research model and an empirical study," Comput. Human Behav., vol. 66, pp. 256–272, 2017, doi: 10.1016/j.chb.2016.09.054.

[34] E. W. Ford, B. W. Hesse, and T. R. Huerta, "Personal Health Record Use in the United States: Forecasting Future Adoption Levels," J. Med. Internet Res., vol. 18, no. 3, p. e73, 2016, doi: 10.2196/jmir.4973.

[35] M. P. Gagnon et al., "Adoption of electronic personal health records in Canada: Perceptions of stakeholders," Int. J. Heal. Policy Manag., vol. 5, no. 7, pp. 425–433, 2016, doi: 10.15171/ijhpm.2016.36.

[36] P. Vezyridis and S. Timmons, "On the adoption of personal health records: some problematic issues for patient empowerment," Ethics Inf. Technol., vol. 17, no. 2, pp. 113–124, 2015, doi: 10.1007/s10676-015-9365-x.

[37] J. Studeny and A. Coustasse, "Personal health records: is rapid adoption hindering interoperability?," Perspect. Health Inf. Manag., vol. 11, 2014.

[38] C. Widmer, J. P. Deshazo, J. Bodurtha, J. Quillin, and H. Creswick, "Genetic counselors' current use of personal health records-based family histories in genetic clinics and considerations for their future adoption," J. Genet. Couns., vol. 22, no. 3, pp. 384–392, 2013, doi: 10.1007/s10897-012-9557-z.

[39] M. D. Logue and J. A. Effken, "Validating the personal health records adoption model using a modified e-Delphi," J. Adv. Nurs., vol. 69, no. 3, pp. 685–696, 2013, doi: 10.1111/j.1365-2648.2012.06056.x.

[40] J. M. Butler et al., "Understanding adoption of a personal health record in rural health care clinics: revealing barriers and facilitators of adoption including attributions about potential patient portal users and self-reported characteristics of early adopting users.," AMIA Annu. Symp. Proc., vol. 2013, pp. 152–161, 2013.

[41] V. M. Sue, M. T. Griffin, and J. Y. Allen, "Beyond adoption: Individual differences in the use of personal health record features in an integrated healthcare organization," Int. J. Biomed. Eng. Technol., vol. 11, no. 3, pp. 252–269, 2013, doi: 10.1504/IJBET.2013.055375.

[42] M. D. Logue and J. A. Effken, "An exploratory study of the personal health records adoption model in the older adult with chronic illness,"

Inform. Prim. Care, vol. 20, no. 3, pp. 151–169, 2013, doi: 10.14236/jhi.v20i3.21.

[43] W.-S. Jian et al., "Factors influencing consumer adoption of USB-based Personal Health Records in Taiwan," BMC Health Serv. Res., vol. 12, no. 1, 2012, doi: 10.1186/1472-6963-12-277.

[44] M. D. Logue and J. A. Effken, "Modeling factors that influence personal health records adoption," CIN - Comput. Informatics Nurs., vol. 30, no. 7, pp. 354–362, 2012, doi: 10.1097/NXN.0b013e3182510717.

[45] J. Curtis, S. Cheng, K. Rose, and O. Tsai, "Promoting adoption, usability, and research for personal health records in Canada: The MyChart experience," Healthc. Manag. Forum, vol. 24, no. 3, pp. 149–154, 2011, doi: 10.1016/j.hcmf.2011.07.004.

[46] P. Rudd and T. Frei, "How personal is the personal health record?," Arch. Intern. Med., vol. 171, no. 6, pp. 575–576, 2011, doi: 10.1001/archinternmed.2011.35.

[47] C. K. Yamin et al., "The digital divide in adoption and use of a personal health record," Arch. Intern. Med., vol. 171, no. 6, pp. 568–574, 2011, doi: 10.1001/archinternmed.2011.34.

[48] T. Greenhalgh, S. Hinder, K. Stramer, T. Bratan, and J. Russell, "Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace," BMJ-BRITISH Med. J., vol. 341, Nov. 2010, doi: 10.1136/bmj.c5814.

[49] M. S. Raisinghani and E. Young, "Personal health records: Key adoption issues and implications for management," Int. J. Electron. Healthc., vol. 4, no. 1, pp. 67–77, 2008, doi: 10.1504/IJEH.2008.018921.

[50] P. C. Tang, J. S. Ash, D. W. Bates, J. M. Overhage, and D. Z. Sands, "Personal health records: Definitions, benefits, and strategies for overcoming barriers to adoption," J. Am. Med. Informatics Assoc., vol. 13, no. 2, pp. 121–126, 2006, doi: 10.1197/jamia.M2025.

[51] B. Heise, C. Asche, and L. Oderda, "RAISE (Rapid Access Integrating Safer Entry) for the Elderly: Readiness of Older Adults to Adopt a Universal Serial Bus Personal Health Record for Medication Reconciliation," Ageing Int., vol. 36, no. 2, pp. 295–302, 2011, doi: 10.1007/s12126-010-9091-y.

[52] E. Andrikopoulou, P. J. Scott, and H. Herrera, "Mixed methods protocol for a realist evaluation of electronic personal health records design features and use to support medication adherence (ePHRma)," BMJ Heal. Care Informatics, vol. 27, no. 1, 2020, doi: 10.1136/bmjhci-2019-100046.

[53] R. Jahn, S. Ziegler, S. Nöst, S. C. Gewalt, C. Straßner, and K. Bozorgmehr, "Early evaluation of experiences of health care providers in reception centers with a patient-held personal health record for asylum seekers: A multi-sited qualitative study in a German federal state," Global. Health, vol. 14, no. 1, 2018, doi: 10.1186/s12992-018-0394-1.

[54] A. Azizi, R. Aboutorabi, Z. Mazloum-Khorasani, M. Afzal-Aghaea, and M. Tara, "Development, Validation, and Evaluation of Web-Based Iranian Diabetic Personal Health Record: Rationale for and Protocol of a Randomized Controlled Trial.," JMIR Res. Protoc., vol. 5, no. 1, p. e39, 2016, doi: 10.2196/resprot.5201.

[55] I. Genitsaridi, H. Kondylakis, L. Koumakis, K. Marias, and M. Tsiknakis, "Evaluation of personal health record systems through the lenses of EC research projects," Comput. Biol. Med., vol. 59, pp. 175–185, 2015, doi: https://doi.org/10.1016/j.compbiomed.2013.11.004.

[56] M. Price, P. Bellwood, and I. Davies, Using Usability Evaluation to Inform Alberta's Personal Health Record Design, vol. 208. 2015. doi: 10.3233/978-1-61499-488-6-314.

[57] B. Sheehan and R. J. Lucero, "Initial Usability and Feasibility Evaluation of a Personal Health Record-Based Self-Management System for Older Adults," eGEMs (Generating Evid. Methods to Improv. patient outcomes), vol. 3, no. 2, p. 3, 2015, doi: 10.13063/2327-9214.1152.

[58] T. Beyan and Y. A. Son, "Incorporation of Personal Single Nucleotide Polymorphism (SNP) Data into a National Level Electronic Health Record for Disease Risk Assessment, Part 3: An Evaluation of SNP Incorporated National Health Information System of Turkey for Prostate Cancer," JMIR Med. INFORMATICS, vol. 2, no. 2, pp. 122–137, 2014, doi: 10.2196/medinform.3560.

[59] A. Sunyaev, "Evaluation of Microsoft HealthVault and Google Health personal health records," Health Technol. (Berl)., vol. 3, no. 1, pp. 3–10, 2013, doi: 10.1007/s12553-013-0049-4.

[60] D. K. McInnes et al., "Development and evaluation of an internet and personal health record training program for low-income patients with HIV or hepatitis C," Med. Care, vol. 51, no. 3 SUPPL. 1, 2013, doi: 10.1097/MLR.0b013e31827808bf.

[61] L. Chiche, A. Brescianini, J. Mancini, H. Servy, and J.-M. Durand, "Evaluation of a prototype electronic personal health record for patients with idiopathic thrombocytopenic purpura," Patient Prefer. Adherence, vol. 6, pp. 725–734, 2012, doi: 10.2147/PPA.S36320.

[62] N. Segall et al., "Usability evaluation of a personal health record.," AMIA Annu. Symp. Proc., vol. 2011, pp. 1233–1242, 2011.

[63] J. Popkin et al., "The eFOSTr PROJECT: Design, Implementation and Evaluation of a Web-based Personal Health Record to Support Health Professionals and Families of Children Undergoing Transplants," in ADVANCES IN INFORMATION TECHNOLOGY AND COMMUNICATION IN HEALTH, 2009, vol. 143, pp. 358–363. doi: 10.3233/978-1-58603-979-0-358.

[64] K. G. Charters and K. Nazi, "Personal health record evaluation: my HealtheVet and RE-AIM.," AMIA Annu. Symp. Proc., p. 899, 2007.

[65] K. T. Win, "Web-based personal health record systems evaluation," Int. J. Healthc. Technol. Manag., vol. 7, no. 3–4, pp. 208–217, 2006, doi: 10.1504/ijhtm.2006.008432.

[66] M. I. I. Kim and K. B. B. Johnson, "Personal health records: Evaluation of functionality and utility," J. Am. Med. INFORMATICS Assoc., vol. 9, no. 2, pp. 171–180, 2002, doi: 10.1197/jamia.M0978.

[67] J. Feather, C. A. A. MISSELBROOK, P. Zipchen, and V. L. L. Matthews, "EVALUATION OF A PERSONAL HEALTH RECORD GIVEN TO NEWBORNS IN SASKATOON," Can. J. PUBLIC Heal. Can. SANTE PUBLIQUE, vol. 78, no. 5, pp. 350–351, 1987.

[68] S. O'Flaherty, E. Jandera, J. Llewellyn, and M. Wall, "Personal health records: An evaluation," Arch. Dis. Child., vol. 62, no. 11, pp. 1152–1155, 1987, doi: 10.1136/adc.62.11.1152.

[69] V. Venkatesh, T. A. Sykes, and X. Zhang, "'Just what the doctor ordered': A revised UTAUT for EMR system adoption and use by doctors," Proc. Annu. Hawaii Int. Conf. Syst. Sci., pp. 1–10, 2011, doi: 10.1109/HICSS.2011.1.

[70] C. B. Jamil Razmak, "Using the technology acceptance model to predict patient attitude toward personal health records in regional communities," A Companion to Philos. Technol., pp. 227–231, 2018, doi: 10.1002/9781444310795.ch41.

[71] J. A. Maxwell, Qualitative Research Design : An Interactive Approach, 3rd ed. SAGE, 2013.

Appendix. I. List of Journals with Methodology Information for Analysis

Journal (N=16)	**Country**	**Period of Study**		**Respondent**				**Methodology**			**Evaluation Model**		
		BFR	**AFT**	**PB**	**P**	**O**	**HP**	**QTY**	**QLY**	**MM**	**TAM**	**UTAUT**	**SDT**
ePHR Acceptance Study (n=6)													
[19]	Saudi Arabia	✓		✓				✓				✓	✓
[17]	Korea	✓					✓	✓				✓	✓
[21]	Canada		✓		✓		✓	✓			✓		✓
[22]	USA		✓				✓	✓			✓		✓
[23]	Taiwan		✓		✓			✓			✓		✓
[24]	USA	✓			✓			✓			✓		
ePHR Adoption Study (n=8)													
[27]	Iran	✓			✓			✓			✓		✓
[30]	UK		✓		✓					✓			✓
[31]	Taiwan	✓		✓				✓			✓		✓
[33]	North America	✓			✓			✓					✓
[32]	Canada	✓		✓				✓			✓		✓
[35]	Canada	✓			✓	✓	✓		✓				✓
[38]	USA		✓		✓					✓			✓
[43]	Taiwan		✓		✓			✓			✓		✓
ePHR Evaluation Study (n=2)													
[53]	German	✓	✓				✓		✓				✓
[60]	USA		✓		✓					✓			✓
% of sub theme		**56**	**50**	**19**	**62**	**6**	**31**	**69**	**12**	**19**	**50**	**12**	**94**

Note:

Period of Study : BFR=Before ePHR Implementation; AFT=After ePHR Implementation

Respondent : PB=Public, P=Patient, O=Organization, HP=Health Professional

Methodology : QTY=Quantitative, QLY=Qualitative, MM=Mix Method

Evaluation Model : UTAUT=Unified Theory Of Acceptance And Use Of Technology, TAM=Technology Acceptance Model, SDT=Self-Determination Theory

Appendix. II. List of Journals with Results and Recommendations for Analysis

Result & Suggestion **Journal (N=16)**	**Result**			**Suggestion**							
	All IV effect DV	**Some IV not effect DV**	**Number of theme identified**	**Increase ePHR awareness or education among stakeholder**	**Conduct sufficient study to identify stakeholders' perspective and need before ePHR development/ implementation**	**Design ePHR application acording to stakeholder need and concern**	**The health proffesional should play a central role to improve utility and consequently the adoption of the ePHR**	**Policymakers, and health-care providers must pay additional attention to increasing individuals conviction and confidence in using the ePHR**	**Healthcare policy-makers, physicians, and developers must consider actions to improve the usability of ePHR in the future**	**Some legal and ethical issues also need to be considered for ePHR adoption**	**Improve the quality of the physician-patient relationship**
[19]	✓			✓	✓						
[17]	✓					✓					
[27]	✓					✓					
[30]	✓		3				✓				
[31]	✓			✓		✓		✓			
[70]	✓			✓			✓		✓		

[53]			4	✓			✓				
[32]		✓				✓	✓	✓			
[33]	✓			✓							
[35]			5	✓						✓	
[22]	✓			✓			✓				
[23]	✓			✓		✓					✓
[60]			3	✓							
[24]	✓						✓				
[38]	✓		4				✓				
[43]				✓			✓	✓			
Percentage of sub theme	**89%**	**6%**	**NA**	**62%**	**6%**	**31%**	**50%**	**19%**	**6%**	**6%**	**6%**

Appendix. III. SUMMARY OF FINDINGS

No.	Objective of Study	Category	Findings
1.	Scope of ePHR Implementation Study	• Study Time Category ◦ Before the implementation of ePHR ◦ After ePHR Implementation • Respondent Category ◦ Patients ◦ Public ◦ Organization ◦ Health Professionals	• 9/16 studies (56%) were conducted before ePHR was implemented • 8/16 studies (50%) were conducted after ePHR was implemented. • 10/16 studies (62%) were conducted on patients • 5/16 studies (31%) were conducted on health professionals • 3/16 studies (19%) were conducted on the general public • 1/16 studies (6%) were conducted on management staff. • 1/16 studies combined 3 groups of respondents (patients, health professionals and management staff) conducted prior to ePHR implementation. • 3 studies involving the public were evaluated before the ePHR was implemented • 6 of the 10 studies involving patients were evaluated after ePHR implementation
2.	Metodology of ePHR Implementation Study	• Methodology Category ◦ Quantitative ◦ Qualitative ◦ Mixed Methods	• 11 studies (69%) were conducted using quantitative methods • 3 studies (19%) were conducted with mixed methods • 2 studies (12%) were conducted with qualitative methods • 7 out of 9 studies (78%) conducted before the implementation of ePHR, were carried out quantitatively. • 7 out of 9 (78%) studies conducted before the implementation of ePHR, were carried out qualitatively. • 4 out of 8 studies (50%) conducted after the implementation of ePHR, were carried out quantitatively. • While 3 studies are mixed methods and the remaining 1 study is carried out qualitatively.
3.	Assessment Model of ePHR Implementation Study	• Assessment Model Category ◦ Technology Acceptance Model (TAM) ◦ Unified Theory of Acceptance and Use of Technology (UTAUT) ◦ Self Determination Theory (SDT)	• 8 studies (50%) used TAM • 2 studies (12%) used UTAUT • 15 studies (94%) established additional variables (SDT) ◦ 7 out of 15 studies (47%) combined SDT with TAM ◦ 2 out of 15 studies (13%) combined SDT with UTAUT ◦ 6 out of 15 studies (40%) used SDT without TAM / UTAUT combination.

Contribution of Experience in the Acquisition of Physical Sciences Case of High School Students Qualifying

Zineb Azar[1], Oussama Dardary[2], Jabran Daaif[3], Malika Tridane[4], Said Benmokhtar[5], Said Belaaouad[6]
Department of Chemistry, Laboratory of Physical Chemistry of Materials
Faculty of Sciences Ben M'Sick Hassan II University of Casablanca, Casablanca, Morocco[1, 3, 5, 6]
Department of Sciences, ENS Meknes, Moulay Ismail University, Meknes, Morocco[2]
Regional Center of Education and Formation in Professions. Boulevard Bir-Anzarane Anfa, Casablanca, Morocco[4]

***Abstract*—This work deals with the importance of experimental practice in the teaching of physical sciences. By practice, we rarely mean practical work carried out in the laboratory. We examined the relationship between students' knowledge of physical science and how practice may or may not help them understand chemical and/or physical concepts. What emerges from a survey distributed to our students is that they are very favorable of the use of the practice. The problem posed in this work consists in identifying the impact of experiments on the acquisition of knowledge and in responding to the problems of learning through experience in the short and medium term. The analysis of the answers allowed us to conclude that the experiment in class, by the teacher, helps to understand the physical and chemical phenomena and can be done before or after the study of the theory. The length and difficulty of practical work sometimes worry students, trying to follow the protocol step by step. This fact underscores the importance of clarity of purpose, through which students can be guided toward questioning what is expected of them, such as knowing how their knowledge has increased.**

Keywords—Experimental practice; physical sciences; students; practical Work PW

I. Introduction

A. Contexte and Problematic

Support for the conduct of experiments and the manipulation of devices by high school students qualifying during practical work has been a long-standing need for teaching the physical sciences: “Physics and chemistry are experimental sciences must be taught as such” [1], could be read in recent official instructions.

From an educational point of view, the goal of the plan is for students to question themselves, act logically and communicate, and build their learning by being representative of scientific activities [2]. Numerous dissertations highlight the advantages of experience in terms of developing transversal skills and increasing the attractiveness of students for the physical sciences, but few are interested in its role in the acquisition of short or medium scientific knowledge term in students. For this reason, it seems interesting to us to study the impact on the students of the manipulation during the experimental phases on the acquisition of scientific knowledge: the realization of experiments allows the students to acquire better knowledge in the short and medium-term?

School instructions regularly evoke the links between theory and experimental practice which plays various roles [3], for example when they recall that experimental activities "constitute the very foundation of physics and chemistry" [4].

There is a large debate around experimental practice to which several researchers have tried to answer. Should theory be taught before practice, or should theory be based on practice?

For Duhem (2016) [5], to say that the theory is built by relying directly on the facts is a mistake, since the theory is not based on experience; it is rather controlled by experience. In other words [6], the physical theory does not start from the experimental facts; it seeks what are the fundamental properties that must be attributed to things and the relations that must be established between the changes of these properties in order to be able to deduce relations from them equivalent to those given by observation.

If we say that theory precedes experience, the best example to cite is that of general relativity proposed by Einstein in 1915. A beautiful, revolutionary theory that nothing could prove at the time. The first experimental verification did not come until 1919 with the Eddington experiment. On the other hand, if we say that experience precedes theory, we have the example of King Frederick II of Denmark who had funded an astronomical research center in the 16th century, which enabled unpublished and precise observations made by Tycho Brahe.

And it is thanks to these observations that Kepler was able to propose his laws on the movement of stars [7].

Constantly concerned about "how" to improve our classroom practice, we have researched, at various levels of our professional curriculum, the difficulties experienced by students in learning physical sciences. We sought, among other things, to know the conceptions of Moroccan high school students qualifying on the role of experience in teaching/learning in physical sciences.

To answer this problem, we will first look at the place of experience in the teaching of physical sciences and then, secondly, we will discuss the methodology used. We will present and analyze the results collected, in order to answer the questions of the problem.

B. Place of Experience in Teaching Physical Sciences

Experimental activities allow the learner to practice the experimental process, criticize, formulate hypotheses, design experiments, to interpret results. We consider that the experimentation helps in a good understanding of the concepts in physical sciences or initiates the construction of knowledge and know-how among the students.

The practical experience can thus be used to consolidate the acquired knowledge and to fill the various gaps of the theoretical course, in order to avoid any type of failure.

Different approaches combine to identify the real factors of failure, by pointing out the gap between saying and doing, between concepts and their uses [8].

Other factors can thus directly influence the decision to act towards such an educational system or a style of learning, in particular in practical work: (a) self-confidence, (b) the need to refer to others, (c) the need to use one's experience, and finally, (d) instilling the training model [9].

The learner profile defined by the program requires that the learner be able to carry out simple scientific experiments capable of highlighting certain facts and detecting the causes. The learner must therefore handle, observe [10].

Experimental activities are carried out in three forms [11]: The first form called "Practical Work" (PW) is a session devoted to manipulations made by students outside the class and guided by the teacher: verification of law, determination of a physical quantity, preparation of solutions, etc.

The other two consist of integrated activities during the course:

- In "course experiments", manipulations are usually done by the teacher, but the teacher sometimes allows the students to manipulate.
- In PW - courses, the manipulations are done primarily by the students.

Putting students in different learning situations allows them to diversify their approaches to a concept. In [12], the more they are confronted with different situations, the more they can discuss this concept and the easier it will be for them to memorize a given concept.

Other experimental learning situations exist, namely the integration of virtual environments for the simulation of various practical works, such as in physics-chemistry [13, 14, 15].

The development of the scientific spirit among learners has thus become more than ever a major objective of science education [16].

The steps to carry out practical work are specified by Joris Deguet and Guillaume Piolle, research professors in their research who specified that the method for doing a practical work takes place schematically in four phases represented in Table I.

TABLE I. PHASES FOR CARRYING OUT PRACTICAL WORK

Phases for carrying out practical work			
Preparation of the experimental protocol	**Carrying out the experiment**	**Collection of results**	**Interpretation and conclusion**
It is provided in the subject of the practical work, and it can be recalled in the report to make it easier to read, or to specify the stage that the student is describing.	The actual manipulations. They follow in practical sessions in the laboratory.	Detail all the experimental results collected during the lab, as well as their precise context. The results are linked to the protocol, to the experiment, and await interpretation, and they do not in themselves constitute a concluding element in a practical report.	This is the key phase of the experimental method, the most important of the report. At this stage, conclusions are drawn, and above all variations in results when the experimental protocol is modified. It is at this stage, by the quality of the writing, that the understanding of the problem is assessed.

C. Obstacles to the Physical Science Experiment

On the practical level, obstacles sometimes stand in the way of the experience, in the case of the educational actors questioned, the lack of equipment and chemicals (silver nitrate for example), the lack of a preparer in the laboratory, and the time allotted for finding equipment and preparing (rehearsing) pose a problem for subject teachers. In [17], the creation of virtual modeling environments can be a good alternative to solve these obstacles.

II. MATERIALS AND METHODS

A. Research Approach

We opted for a quantitative study based on the exploration and description of the data collected, as well as the comparison with related studies for discussion. The objective of our chosen approach is to get a set of concrete ideas of our research problem and objective.

We chose to use a quantitative method instead of a qualitative one, because:

- it reassured the interviewee and increased the response rate;
- it clarifies ambiguous answers;
- it allows us to transmit the questionnaire to the person;
- it allows the questionnaire to be passed on to the person who is competent to answer it; and
- suitable for questionnaires with "drawers".

B. Sample

The surveys are conducted, during the 2018-2019 school year, in two provincial directorates of the Ministry of Education and National Training; Sidi Bernoussi and Moulay Rachid belonging to the Casablanca-Settat regional. The choice of these directorates is related to reasons of facilitation of the research process.

Our sample consists of n = 300 qualifying secondary school students (common core and first year of the baccalaureate), 94% of the students responded because the questionnaire was given to them in class and they had time to answer it. Indeed, simple random sampling (SRS) was used in our study.

C. Method of Data Collection

We opted for the anonymous questionnaire as our data collection instrument. The questionnaire items were divided into four parts:

- Organization of the practical sessions: determine if the organization of the practical sessions and the objectives are clear enough.
- Theory and experience: determine if one is separable from the other.
- Role of experience: determine how the student perceives learning through experience.
- Progress of the practical work sessions: determine how the pupils perceive the availability of the teacher.

And finally, a section where students could freely answer three open questions:

- In your opinion the practical sessions allow us to acquire new knowledge and skills?
- In your opinion, what are the strengths and weaknesses of TPs?
- Did the practical sessions help you understand the course in physical sciences?

III. Results and Discussion

After the collection of data by the questionnaire, we proceeded to a statistical treatment using Excel software. The results were shown in the various figures below.

A. Organization of Practical Work Sessions

Fig. 1 gives the results of the two questions relating to the teacher's explanation and the clarity of the objectives of the practical sessions as answered by the respondents:

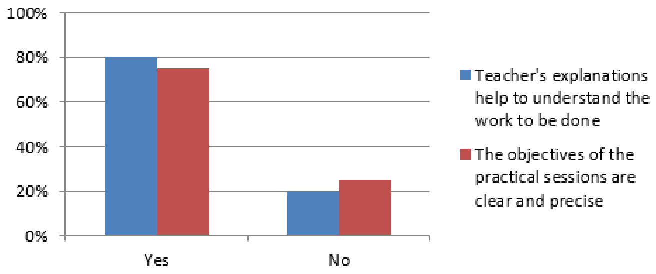


Fig. 1. Organization of Practical Work Sessions.

It turns out that: For 80% of the students questioned, the teacher's explanations help to better understand what is expected of them, and 75% declare the clarity of the objectives of the practical sessions.

Work on effective teaching practices [18-29], show that the teaching activities of efficient and equitable teachers are characterized by the principle of clarity, clarity in the presentation of tasks and instructions to be performed by students, clarity in the general organization of activities teaching. The explanation of concepts and concepts is done simply and logically.

What we noticed during the exercise of this profession of professor of physical sciences, is that, even if it is certain that the pupils should not be left in ignorance of the objective of practical work, they have a relative lack of autonomy. Indeed, as soon as part of the work is left to their judgment, we can say that they do not feel comfortable and ask a lot of questions.

Among the stated goals of good practical work in numerous publications are the familiarization of students with the scientific process [30]; [31], cognitive and psychomotor skills (Galloway et al., 2016) [32] and a better understanding of how science works [30, 31], without forgetting that there is a strong link between a learner's motivation and his results (Deci, Vallerand, Pelletier, Ryan, 1991) [33].

B. Theory and Experience

In Fig. 2, we have presented the results of the student survey concerning their opinion on the theory of experience.

87% of students believe that theory must precede experience, which again shows a relative fear of the student facing the unknown, a lack of self-confidence and autonomy.

In conclusion, for the pupils, it is useless to see the experiments before the theory; consequently, they prefer to already know the subject on which they must concentrate.

According to the literature many experiments were carried out without being guided by a well-established theory but were instead undertaken to explore new areas. Let's look at some illuminating examples: The development of electromagnetism is a good example: many electrical phenomena were discovered by physicists such as Charles Dufay, André-Marie Ampère, and Michael Faraday in the 18th and 19th centuries through experiments that did not have been guided by no developed theory of electricity.

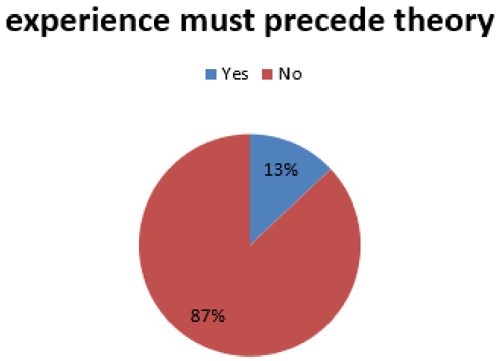


Fig. 2. Theory Vice Versa Experience.

Another example to cite: is the constant and laborious experimental effort that characterized the field of particle physics, especially the fundamental constituents of the proton and neutron, which were discovered through several exploratory experiments at the end of the 1960s to the Stanford Linear Accelerator.

For the second question of this item, we find in Fig. 3 that Almost 89% of students think that experience helps them understand the theory. When we analyzed the results, we could read that assisting or experimenting helps them build an image of the theory and keep it in mind.

The practical work during which the student is supposed to think, act and manipulate is always preceded by theory, where the knowledge at stake in these labs is already in place, which leads to reinforcing the lack of autonomy among the students of the second qualifying.

These results do not mean that experience should not have a close relationship with theory; in the end, there is only one physics, and it must ensure unity [7]. So, we can deduce that it is up to the teachers to clarify to the students the importance and the complementarity between the theory and the experiment whatever the first one.

For the last question in this section, we find in Fig. 4 that 70% of students think that the physical sciences cannot be taught without carrying out the experiments and also that the description of the experiment alone is not enough to understand the theory (76%). On the other hand, they need theory and practice to understand physical and chemical concepts.

The teachers want the PW to develop scientific attitudes as well as a creative and critical dimension in the pupils, the latter for their part, are much more reserved about the achievement of the objectives characterizing scientific attitudes.

They perceive the practical work under an aspect linked to the manipulative techniques and the illustration of the course [34]. On the other hand, we can find a part of the teachers who declare that the physical sciences can be taught without experimenting, in particular, at the forefront of national concerns and priorities in the discourse of the main political actors in Morocco [35].

C. *Role of Experience*

76% of the pupils think that the pupil learns more from the experience than he did than that of the teacher, 77% prefer the experiences which were made in the group. All this is linked to the previous question, which asks them if practical work helps them to better understand the physical sciences (Fig. 5).

The learning of manipulation techniques is facilitated, according to the literature and the teachers, by precise descriptive protocols, supplemented by illustrations and diagrams of the technique to be understood for the labs carried out by students in groups [36].

The teachers, for their part, emphasize their role to play in this learning as a student guide, in particular by validating or correcting the student's actions during the practical work [37].

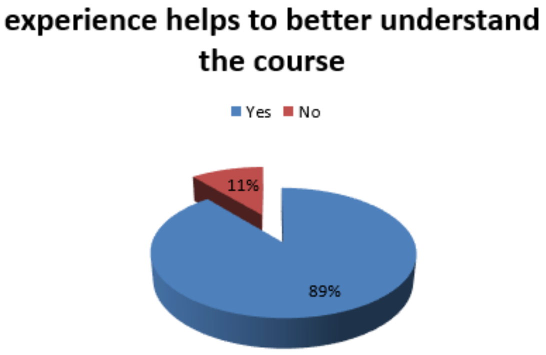


Fig. 3. Theory and Experience.

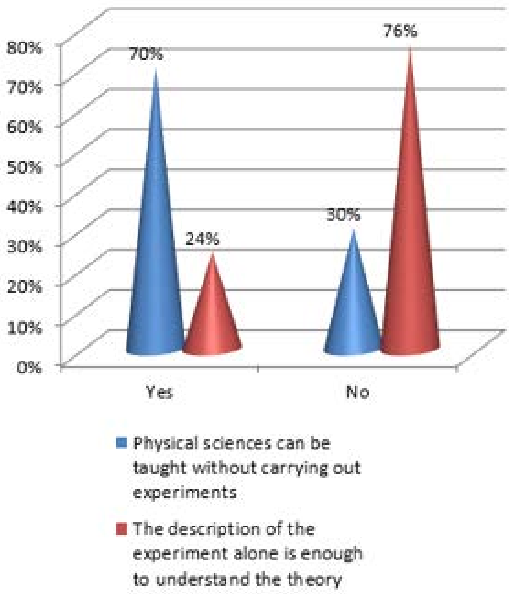


Fig. 4. Physical Sciences can be Taught without Experience.

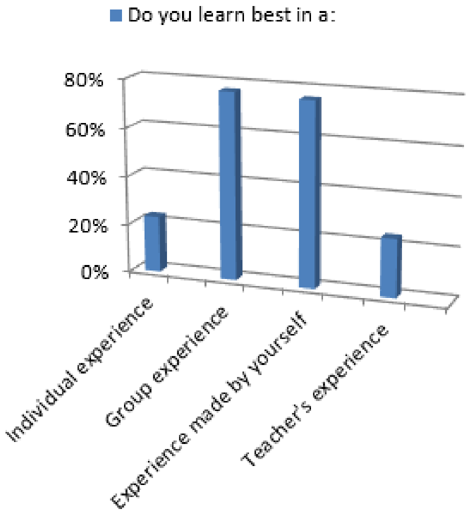


Fig. 5. Types of Experience that Students Learn Best.

D. Progress of the Practical Work Sessions

For students, the lesson must be flexible enough to integrate the discussion on a topical subject if the opportunity arises; for this, we asked the student's opinions on the course of the practical sessions (Fig. 6).

It appears that:

- The majority of the students questioned affirm that the teacher helps to build knowledge and overcome the difficulties encountered in understanding physical and chemical concepts.
- Most students are very interested in experimental activities; with the majority finding that the teacher helps them better understand the course.
- The results of this document have shown that the use of experience (under the guidance of the teacher), as an educational means in the teaching of the physical sciences and plays an important role among learners in the plan of acquisition of correct knowledge and concepts related to the program.

The teachers, for their part, emphasize their role to play in this learning as a student guide, in particular by validating or correcting the student's actions during the session.

For the last part concerning the three open questions, is not discussed in this report. The reason is that the majority of students surveyed did not answer these questions.

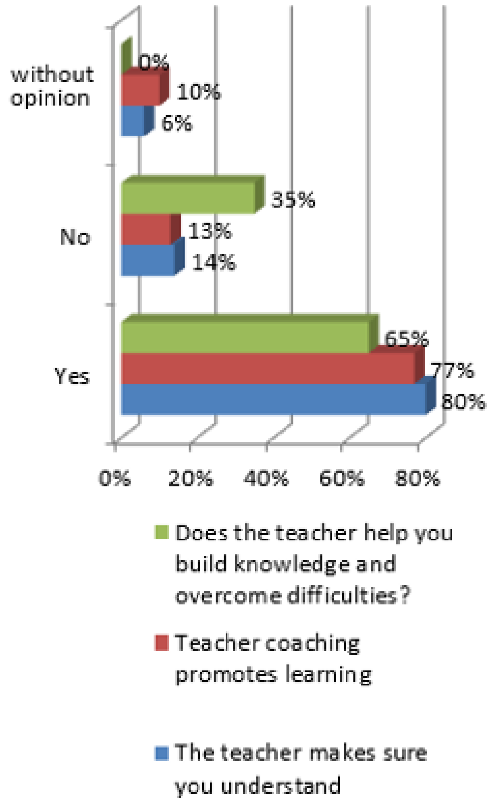


Fig. 6. Role of the Teacher in the Practical Work Sessions.

IV. Conclusion

Our work aimed to explore the role played by experience from the students' point of view. The results obtained according to the statistics were favorable and benevolent and go to achieve the objective of the research and respond to the problem raised. The results that we can draw from our brief experience allowed us to raise many points of interest in our opinion:

- The importance of the physical sciences comes through experimentation, whether it is a demonstration in the classroom or the form of practical work, carried out in groups by the students themselves.
- The classroom teacher's experiment helps to understand physical and chemical phenomena and can be done before or after studying the theory.
- Try to vary as much as possible in the practical work sessions, showing the theory before the experiment or letting them discover the theory. We could see that they preferred to know the theory before applying it to the labs.
- Practical work PW helps students to visualize the concepts taught, explore and develop their critical thinking skills.

We can recommend to information and communication technologies, in particular carrying out practical experiments using digital tools in environments which depend on certain factors which seem to influence these practices.

Indeed, the reflection on the modalities of digital use by students gives rise to several debates [38].

In addition, and a context of virtual experiential learning, several pedagogical approaches can reinforce the playful learning approach, particularly at a distance, which thus represents a powerful vector for promoting socio-professional skills [39].

To this is added the importance of the educational act and the effective role played by the school in the construction of society based on solidarity and equal opportunities in different social and environmental environments [40].

V. Recommendations

For research in didactics, this study can serve as a support for various didactic reflections on experimental practice and in particular, the development of practical work in physical sciences taking into account the opinions of learners. It can allow a reflection on an interaction between pupil and teacher and, between groups of pupils to promote their autonomy and participation in the construction of their knowledge.

An interesting avenue for future research would be to seek to confirm these results by using a larger sample by conducting more similar experiments in other scientific disciplines to be able to draw a more general conclusion.

In addition, we can identify in future research the constraints and obstacles hindering the integration of information and communication technologies for practical laboratory work (lack of training, software, appropriate materials, assistance administrative, etc.).

We can thus take advantage of the learning difficulties often detected when learners do not reach the level required for their age groups. They can take the form of poor motivation, limited memorization or concentration, inability to solve problems, maladaptive social behaviors, etc. [41].

The concepts of evaluation and oral communication are also important, making it possible to gradually improve learners' communication techniques and skills, while maintaining minimum stress and keeping the oral communication project in its persuasive dimension [42].

We can refer to the various works of educational researchers, such as those of Fallon in 2019 who explained that the first research focused on understanding scientific concepts using physical equipment, but technological advances mean now that there are new options for introducing this learning, through the use of technology and simulation.

Indeed, the results indicate that the students developed a solid foundation of procedural knowledge about building different circuits and functional knowledge about circuit components which they applied to different circuit designs.

Adding that teachers need to be very vigilant and work closely with learners, to ensure that accurate understanding is developed [43].

We can thus appeal to the research of Yardley's 2012 [44], research aim to introduce readers to the theories underlying experiential learning, which considers the theoretical basis of experiential learning from a learning perspective social and constructionist and applies it to three stages of medical education: the first experience in the workplace, internships, and residency.

Another study investigated the level of usefulness of e-module based on experiential learning in the physical sciences. The authors worked on Research and Development (R&D). The method used is a descriptive method with data collection instruments, namely the usefulness questionnaire of the electronic module evaluated by the teacher.

The results showed that e-modules based on experiential learning were very useful in learning physics [45].

The results obtained can be exploited in future attempts to ensure better integration of the experience in the Moroccan education system, the main actors must collaborate to overcome the obstacles that prevent this integration, namely: the Ministry of Education and of training, and the teacher.

VI. Limitations

Although this study yielded important results in the quantitative database dealing with the importance of experimental practice in the teaching of the physical sciences, it has certain limitations.

Indeed, the research is limited to a small sample belonging to two provincial directorates of the Ministry of Education and National Training; Sidi Bernoussi and My Rchid belonging to the Casablanca-Settat regional academy in Morocco.

We did not consider the rural environment. This was due to the lack of equipment and resources available to the research team, which forced us to work only in the Casablanca-Settat region.

Another limitation is that we have relied solely on the physical sciences, rather than other important scientific disciplines that are meant to be incorporated and generalized in this study.

Acknowledgment

We warmly thank all the teachers and all the administrative staff of the chemistry department affiliated with the Ben M'Sick Faculty of Sciences of the Hassan II University of Casablanca for their support and help in carrying out this work.

References

[1] KOUELA, P. (2012). Le rôle de l'expérimentation dans l'enseignement de la Chimie au Lycée : cas des réactions chimiques. Université Marien Ngouabi - DEA 2012 ; Ecole Normale Supérieure (République du Congo).

[2] Ajchenbaum-Boffety B., Chevalerais, F., Chomat, A., Desbeaux-Salviat, B., Ernst, S., Jasmin,D., Larcher, C., Renoux, Y., Saltiel, E, Sarmant, J-P. (2000). La main à la pâte et le Plan de rénovation de l'enseignement des sciences et de la technologie à l'école. Guide de découverte. Brochure de l'Académie des sciences. Paris : Service des publications de l'Institut National de Recherche Pédagogique (INRP).

[3] Cariou,J-Y., (2015). Le statut épistémologique de l'expérience dans les nouvelles approches préconisées pour l'enseignement des sciences », RDST, (12), pp. 59-85.

[4] ASTOLFI, J-P., PETERFALVI , B. & VÉRIN ,A. (1998). Comment les enfants apprennent les sciences. Paris : Retz.

[5] DUHEM,P., (2016) .La Théorie physique. Son objet-sa structure, p. 336.

[6] DUHEM,P., (2010). L'expérience comme interprétation des faits dans la " théorie physique ",

[7] Panoutsopoulos,G.,Zimmermann,F. (2019). Which Should Come First in Physics: Theory or Experiment? Scientific american.

[8] Maziane, B. , Bassiri, M. , Benmokhtar, S., Belaaouad , S. (2020). Engineering analysis of teaching practices and learning strategies guided by the principles of Cognitive Psychology and Information technology. International Journal of Advanced Trends in Computer Science and Engineering. (9). pp. 212 - 217.

[9] Z., Habybellah, M. Bassiri, S., Belaaouad, M., Radid, S. Benmokhtar. (2019). Training and Development of Professional Skills: An Analysis of activity in Professional Skills. International Journal of Advanced Trends in Computer Science and Engineering. 8. 2029-2033. 10.30534/ijatcse/2019/28852019.

[10] KOUELA, P. (2012). Le rôle de l'expérimentation dans l'enseignement de la Chimie au Lycée : cas des réactions chimiques. Université Marien Ngouabi - DEA 2012 ; Ecole Normale Supérieure (République du Congo).

[11] Kane, S. (2011). Les pratiques expérimentales au lycée- Regards croisés des enseignants et de leurs élèves. Radisma, (7), pp. 1-26.

[12] Palacio-Quentin, E. (1990). L'éducation cognitive à l'école. European Journal of Psychology of Education, (2), pp .231-242.

[13] Daaif, J., Zerraf, S., Tridane, M., El MahiChbihi, M., Moutaabbid, M., Benmokhtar, S., Belaaouad, S. (2019a). Computer simulations as a complementary educational tool in practical work: Application of monte-carlo simulation to estimate the kinetic parameters for chemical reactions. International Journal of Advanced Trends in Computer Science and Engineering, 8(1.4 S1), pp. 249–254. https://doi.org/10.30534/ijatcse/2019/3881.42019.

[14] Daaif, J., Zerraf, S., Tridane, M., Benmokhtar, S., Belaaouad, S. (2019b). Technological Innovation in Teaching and Research in Chemical Science: Development of a Computer Application for the Simulation of the Practical Works of Crystallography. International

Journal of Recent Technology and Engineering (IJRTE). Volume-8 Issue-3. DOI: 10.35940/ijrte.C4665.098319.

[15] Daaif J., Zerraf S., Tridane M., Benmokhtar S., Belaaouad S. (2019c). Pedagogical engineering to the teaching of the practical experiments of chemistry: Development of an application of three dimensional digital modelling of crystalline structures. Cogent Education, 2019, 6(1), 1708651.

[16] Sall,Ch.T.,Kane,S. (2007). Quand les élèves parlent de l'enseignement de la physique et de la chimie et des pratiques expérimentales au lycée. Sientia paedagogica Experimentalis, pp .1-3.

[17] Daaif, J., Zain, S., Zerraf, S., Tridane, M., Khyati, A., Benmokhtar, S., Belaaouad, S. (2019d). Progress of Digital Learning Resources: Development and Pedagogical Integration of a Virtual Environment Laboratory for the Practical Experiments in Chemistry. International Journal of Innovative Technology and Exploring Engineering (IJITEE). Volume-8 Issue-11. DOI: 10.35940/ijitee.K2403.0981119.

[18] Attali, A. & Bressoux, P. (2002). L'évaluation des pratiques éducatives dans les premiers et seconds degrés. Paris : rapport établi à la demande du Haut conseil de l'évaluation de l'école.

[19] Bressoux, P. (1994). Les recherches sur les effets-écoles et les effets-maîtres. Revue Française de Pédagogie,(108), pp. 91-137.

[20] Bressoux, P. (2000). Pratiques pédagogiques et évaluation des élèves. In A. Van Zanten (dir.), L'école l'état des savoirs pp. 198-207. Paris : Éditions la découverte.

[21] Bressoux, P. (2007). Des compétences à enseigner : quelles « traces » sur les apprentissages des élèves. In M. Bru & L. Talbot (dir.), Des compétences pour enseigner, Entre objets sociaux et objets de recherche, pp. 121-134. Rennes : PUR.

[22] Cusset, P.-Y. (2011). Que disent les recherches sur l' « effet enseignant » ? La note d'analyse, pp. 232.

[23] Dumay, X. & Dupriez, V. (2009). L'efficacité dans l'enseignement. Promesses et zones d'ombre. Bruxelles : De Boeck.

[24] Duru-Bellat, M. & Mingat, A. (1988). Le déroulement de la scolarité au collège : le contexte fait des différences. Revue française de sociologie, 29(4), p. 649-666. DOI : 10.2307/3321516.

[25] Felouzis, G. (1997). L'efficacité des enseignants. Paris : PUF.

[26] Mingat A. (1983). Évaluation analytique d'une action Zone d'Éducation Prioritaire au cours préparatoire. Cahier de l'IREDU, 37.

[27] Mingat A. (1984). Les acquisitions scolaires au CP : l'origine des différences ? Revue Française de Pédagogie, (69), pp. 49-62.

[28] Mingat, A. (1991). Expliquer la variété des acquisitions au cours préparatoire : les rôles de l'enfant, la famille et l'école. Revue Française de Pédagogie, (95), pp. 47-63. DOI : 10.3406/rfp.1991.1355.

[29] Safty, A. (1993). L'enseignement efficace. Théories et pratiques. Québec : Presses de l'Université du Québec.

[30] DeKorver, B. K. et Towns, M. H. (2015). General Chemistry Students' Goals for Chemistry Laboratory Coursework. Journal of Chemical Education, (92), pp. 2031-2037.

[31] Moore, J. W. (2006). Let's Go for an A in Lab. Journal of Chemical Education, (83), pp.519.

[32] Galloway, K. R., Malakpa, Z. et Bretz, S. L. (2016). Investigating Affective Experiences in the Undergraduate Chemistry Laboratory: Students' Perceptions of Control and Responsibility. Journal of Chemical Education, (93), pp. 227-238.

[33] Deci, E. L., Vallerand, R. J., Pelletier, L. G. et Ryan, R. M. (1991). Motivation and Education: The SelfDetermination Perspective. Educational Psychologist, 26(3-4), pp. 325-346.

[34] C. Génin et A. Pellet.(1993),Etudiants et enseignants face aux travaux pratiques de physique en 1ère année de DEUG , Tréma, 3(4) , pp. 93-107.

[35] Azar, Z., Dardary, O., Tridane, M., Benmokhtar, S., & Belaaouad, S. (2021). Bibliographic Analysis on the Financing of Education Reform in Morocco. Iraqi Journal of Science, 276-281.

[36] Domin, D. S. (1999a). A Content Analysis of General Chemistry Laboratory Manuals for Evidence of Higher-Order Cognitive Tasks. Journal of Chemical Education, 76(1), pp.109-111.

[37] Green, W. J. et Elliott, C. (2004). "Prompted" Inquiry-Based Learning in the Introductory Chemistry Laboratory. Journal of Chemical Education, 81(2), pp. 239-241.

[38] FIKRI,R. , BELHABRA, M. , KHOUYA,E. , TRIDANE M. ,& BELAAOUAD, S. (2018). The use of digital modalities among Moroccan PhD students: Case of Hassan II University of Casablanca. ASIA LIFE SCIENCES Supplement, 16(1), pp. 239-245.

[39] BASSIRI,M. , Boulahouajeb, A., AICHI,Y., BELAAOUAD, S. et RADID , M. (2018). Distance learning - A powerful vector of the enhancement of socio-professional competences: Case of the training of contract teachers in Regional Centre of Training and Education, Morocco. ASIA LIFE SCIENCES Supplement, 16(1),pp.11-19.

[40] Zain, S, Zerraf, S, Belaaouad, S, khyati, A. (2021). The Policy of Integration in Developing The Skills of People with Special Needs Through A Professional Didactic Approach "Analytical Field Study on the Reality of Integrated Classes in Azrou City". Iraqi Journal of Science. 356-362. https://doi.org/10.24996/ijs.2021.SI.1.49.

[41] Zain, S., Daaif, J., Zerraf, S., Belaaouad, S., khyati, A. (2019). Technological Implication and Pedagogical Effects of Reading Difficulties on the Linguistic Achievements of the Moroccan Child for Some Primary Schools. International Journal of Innovative Technology and Exploring Engineering (IJITEE). Volume-8, Issue-11. DOI: 10.35940/ijitee.K2404.0981119.

[42] Zerraf, S., Bassiri, M., Zain, S., Tridane, M., Belaaouad. S. (2021). Engineering of An Innovative Oral Scientific Communication Device: Case of the Doctoral Training Context. Iraqi Journal of Science.133-140. https://doi.org/10.24996/ijs.2021.SI.1.18.

[43] Falloon, G. (2019). Using simulations to teach young students science concepts: An Experiential Learning theoretical analysis. Computers & Education. doi:10.1016/j.compedu.2019.03.001.

[44] Sarah Yardley, Pim W. Teunissen & Tim Dornan (2012) Experiential learning: Transforming theory into practice, Medical Teacher, 34:2, 161-164, DOI: 10.3109/0142159X.2012.643264.

[45] Fadieny, N., & Fauzi, A. (2021). Usefulness of E-module Based on Experiential Learning in Physics Learning. International Journal of Progressive Sciences and Technologies, 25(1), 410-414.

Fit-Gap Analysis: Pre-Fit-Gap Analysis Recommendations and Decision Support Model

LAHLOU Imane, MOTAKI Nourredine, SARSRI Driss, L'YARFI Hanane
Nationale School of Applied Sciences of Tangier
Tangier, Morocco

***Abstract*—Enterprise Resource Planning (ERP) system has been defined as a configurable Commercial Off-The-Shelf (COTS) system integrated into multiple business functions. For most companies, adopting ERP has become necessary to maintain market competitiveness. However, ERP implementation is still critical because project success depends on multiple parameters and involves several stakeholders. This article deals with the Fit-Gap analysis stage, which is an essential step in ERP implementation. This study was carried out through a literature review and interviews with experts to gather information and support stakeholders toward a successful Fit-Gap phase. It presents a set of recommendations for clients and consultants to consider before starting the Fit-Gap Analysis phase, and it presents an approach, with a decision support model represented as Business Process Modelling Notation (BPMN) based on several parameters to be used during the Fit-Gap Analysis stage to bridge gaps. The results obtained are intended for clients and consultants to make the most rational decision to bridge gaps based on the recommendations found, the approach and the decision support models presented.**

***Keywords*—ERP systems; misfit; customisation; fit-gap analysis**

I. Introduction

Information systems are at the company's core [1]. Indeed, an information system is composed of workers, processes, systems, applications, databases, and rules. It is accountable for storage and information processing. It provides information to the employees [1]. Successful company management is attainable by providing the right information to the right person. However, an efficient information system helps to have a productive and competitive company. Many enterprises have and use an Enterprise Resource Planning (ERP) system for its managerial characteristics.

ERP is a software package that automates and integrates a company's business processes [2]. Nowadays, Small and Medium-sized (SMEs) and Large Enterprises are adopting these systems for their many organisational advantages.

Scientific research has focused on ERP implementation since it is a significant phase in its life cycle. Researchers have also emphasised some key success factors, given the high degree of failure rate cited in previous research and observed in the field [3] [4] [5] [6].

Several success factors for the implementation project have been mentioned in the literature, among which we found that the ERP was not adapted to the company's business processes. Indeed, the ERP does not always meet the customer's needs. Here, we notice a gap, or what is also known as a misfit, between the system and the company's business process. Therefore, this gap can be bridged in several decision-making [7]:

- Entirely adopt the ERP standards and completely abandon its processes (we called this the vanilla method) through the Business Process of Re-engineering (BPR).
- Waive some of its established requirements.
- Find workarounds to bridge this gap (this may be achieved by introducing other applications or manual work).
- Customise the ERP system to fit the company's processes.

The company should decide the strategy to align the ERP systems with the customer needs and business processes, which will inevitably be a project success factor. Leading and evaluating this decision-making is an interesting aspect to consider. Therefore, it is going to guarantee the project's success. Making the right decision is a key factor to avoid falling into a standard problem of ERP systems implementation i.e., over-customisation.

ERP over-customisation is a classical issue. Several research pathways have approached the issue from different angles to address it, in a theoretical or empirical way. However, the tools and methods that have been developed only propose a partial solution to the problem. Hence, the need to develop a new study includes a solution to solve the entire issue. Even though scientific research has a keen interest in this problem, it is still topical to find companies that do not handle having a system set up correctly to meet their needs. The cause is often because of unsuccessful management of gaps.

This article deals with the Fit-Gap Analysis phase by presenting first, the literature review concerning this phase; Second, recommendations to be taken into consideration before starting the Fit-Gap Analysis stage. And finally, an approach and decision support model to address gaps.

II. Literature Review

To define this study and understand the proposed problem, the following section discusses some points that are directly related to the subject through a literature review i.e., ERP misfit, misfit types, Fit-Gap Analysis, ERP customisation, and synthesis.

A. ERP Misfit

ERP misfit is a derivation of a broader concept called Task Technology Misfit (TTF) to explain the gap between ERP systems in terms of technology and the organisation's requirements [2]. Fig. 1 explains it in an illustration form [8]. Indeed, several studies have used this concept to have improved clarification regarding ERP misfit [9] [3] [10]. Therefore, it would be wise to learn more about the TTF and investigate the topic to have a better understanding of ERP misfit.

The TTF theory shows the degree of Information Technology (IT) capability to support user requirements. It is important to state that information systems can only have positive effects on the system or the user performance if they align with the user requirements in terms of tasks [11]. It is, therefore, clear that TTF is an excellent system performance indicator and even better than user satisfaction or technology use [12].

The projection of the TTF theory on ERP systems is, therefore, being the ERP misfit. To have an efficient ERP and better use of it, it is important to have a high degree of fits and consequently a very low degree of misfits. The primary cause of failure in implementation projects reports this ERP misfit with the customer's requirements [13]. This is a real problem to be taken seriously when implementing an ERP in a company. This is because it could cause much more damage than poor performance or reduced flexibility and agility [14]. It could also lead to large financial losses on the scale of hundreds of millions of dollars, or even near-bankruptcy [15].

B. Misfit Types

Many studies have classified misfits according to several categories and different angles, ranging from broad categories to subcategories. There are perceived misfits and real misfits [16]. It remains significant to distinguish the two misfits to know how to solve them [17].

Perceived misfits are misfits that do not cause any difficulty in the business or taxes. They are usually caused by ERP systems functioning ignorance, unrealistic customer expectations, and resistance to change. No matter how well the ERP technically works, it can still face opposition. So having motivated users and cooperation from the business team is crucial [18].

Real misfits are the true inadequacy of the ERP systems with the customer needs, including imposed misfits and voluntary misfits [15]. The imposed misfits are therefore due to either an obligation set by the authorities or by industrial standards [19], contrary to the voluntary misfits, which are company choices to distinguish itself in the market [7].

There are two other misfit types that can be considered subcategories of the previous types, "deep structures" and "surface structure" [20]. The deep structure type is at the core of misfits. They occur when things, properties, states, or transformations are erroneous or represent deficiencies in the system [21]. In addition, surface misfits can cause mismatches if the software does not offer the proper format for reports, provide the appropriate roles to access information, or have a suitable user interface [7].

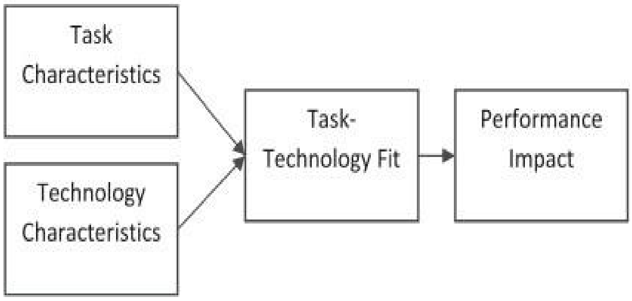


Fig. 1. Task Technology Misfit Diagram.

There is another misfits classification made by Soh et al. by grouping the most common mismatches into three categories: input, processes, and output misfits [21].

Input misfit is the inability of the ERP systems to capture different objects, attributes, or documents in the ERP database [22]. It is a misfit concerning the data entries and the diverse relationships between the entities in the data model [21]. It is natural that the absence of certain elements, in terms of data, will lead to severe software deficiency and will affect its performance [2]. Therefore, input misfits can be included in the deep misfit category [16].

Process misfit refers to, the incompatibility between the business requirements and the ERP systems in terms of required processes and procedures [21]. ERP is supposed to represent the workflow activities of the enterprise as they are in the real world. This inability to model them in the ERP system is a process misfit. Typically, that will affect information quality, presenting the user with poor quality or illogical information [2]. This misfit category is also included in the deep misfit [16].

Output misfit is the incompatibility between the ERP systems and the business requirements, in terms of information format presentation, interface content, or reports. The information which is poorly presented to the user will take an extended amount of time to be understood and could eventually be misinterpreted [23]. Therefore, output misfits can be included in the surface misfits [16].

Another category can be added to the three other types. Established by Soh et al, it is the latent structure [20] containing three additional categories proposed by D.M. Strong and O. Volkoff, i.e., role misfit , control misfit, and organisational culture misfit. In this article, only input, process, and output misfits are treated. Fig. 2 represents misfit types according to Eli et al., Joost A. A. van Beijsterveld and Willem J. H. van Groenendaal and Tan Shiang-Yen et al. [7] [16] [2].

C. Fit-Gap Analysis

Fit-Gap Analysis is a critical phase in the ERP life cycle. It is used for ERP systems selection and to generate inputs for the design phase [24]. Therefore, we note high-level Fit-Gap analysis in ERP pre-implementation and detailed Fit-Gap Analysis during implementation projects [25]. Fit-Gap Analysis is conducted by several stakeholders: ERP vendors, consultants in charge of the implementation project, and customers who will implement the ERP systems [26].

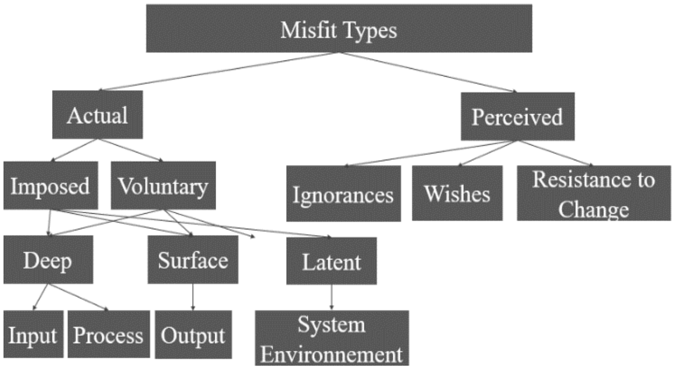


Fig. 2. Misfits Type.

In high-Level Fit-Gap Analysis, the aim is to select the most suitable ERP systems for the company [8]. It is based on a fit between the customer requirements and the ERP systems capability. It is, therefore, a matter of minimising the gap between the ERP systems and the customer needs by selecting the ERP that best meets these criteria. The right choice will increase the chances of successful project implementation.

The high-Level Fit-Gap Analysis can be done by following these steps [11]:

- a review of customers' business needs and ERP systems capabilities.
- the selection of a comparison approach.
- a comparison of business needs with ERP systems capabilities and the documentation of fits and gaps.

Once the most suitable ERP is selected, the company can move on to the next phase, which is the detailed Fit-Gap Analysis. Indeed, this phase is critical. It is about collecting all the misfits and fits between the ERP and the customer requirements. As soon as the set of gaps is determined, it should be resolved by a successful implementation of the project [24]. The gap resolution can be done either by customising the ERP system or by adjusting to business processes. The detailed Fit-Gap Analysis is therefore used to determine the degree of change that needs to be made to both the softwares and the company [27].

D. ERP Customisation

Customisation is a term used to refer to ERP modification to align with customer needs [28] [29] [30]. The customer specifies his business requirements. If the ERP meets them, there will be no need to customise the system. Otherwise, the client is left with several alternatives, among them, ERP customisation [31]. To make such a decision, several parameters come into play. We often find the cost-benefit aspect as key [32]. However, some important parameters are not always taken into consideration by stakeholders. For instance, S. Koch and K. Mitteregger insist on the study maintenance and its costs to decide whether to customise ERP systems or not [33].

Despite the many benefits that ERP customisation can bring, such as increasing the local efficiency of the company [34], and user satisfaction [35], most scientific research remains categorical about minimising customisation to achieve a successful implementation project [36] [37]. Indeed, customisation increases the duration of the implementation project. It can generate bugs and update issues. One can even say that it can jeopardise the many benefits that ERP can bring [2] [38]. However, it is still usually necessary. The question one might ask is, what limit must never be exceeded to not sway towards over-customisation?

E. Synthesis

In the literature there are various studies proposing several techniques and methods to provide a solution to ERP misfit, several evaluation methods have been proposed to treat the issue partially.

Numerous studies proposed to study the relationship between ERP customisation and system maintenance. S. Koch and K. Mitteregger conducted a study to evaluate the relationship between customisation degree and support effort [33]. While B. Light presented, two cases, describing customisation performed and the maintenance implications [28].

S. Parthasarathy and S. Sharma realised the research on the customised ERP efficiency. They adopted a quantitative method using Data Envelopment Analysis model [37]. The purpose is to examine customised ERP efficiency and the relationship between customisation degree and ERP package efficiency [37].

Customisation can also impact ERP quality. S. Parthasarathy realised an empirical study based on framework and advanced a set of hypotheses to establish the relationship between the customisation carried out by ERP vendors and ERP resulting quality [31].

Customisation has also to take into account different types of ERP systems. Elin Uppström et al. proposed different options for Cloud ERP and One-premise ERP and discuss the difference between there [39].

C. S.-P. Ng treated the relationship between ERP alignment types, operational characteristics, degree of system use, user satisfaction, and system benefits [35].

Hustad et al. have described the different types of misfits and types of tailoring [7].

Table I (Appendix) is a synthesis of these studies carried out in the previous research about the Fit-Gap Analysis phase, ERP misfit, and ERP customisation.

Based on the research related to ERP misfit, Fit-Gap Analysis, and ERP customisation, it is clear that there is a need for a study that takes into account all of this research to support consultants and clients during the pre-Fit-Gap Analysis phase and the Fit-Gap Analysis.

III. Research Methodology

This study aims to avoid redundant errors in previous projects in the Fit-Gap Analysis phase. A set of recommendations are presented concerning Fit-Gap Analysis. This recommendation should be reviewed before starting this stage. Later in this paper, an approach and a support model are presented as Business Process Modelling Notation (BPMN) to support stakeholders in this phase. These two main objectives can be formulated as three questions:

RQ1: What are the recommendations that should be reviewed before starting the Fit-Gap Analysis phase?

RQ2: What are the parameters to consider during the Fit-Gap Analysis phase?

RQ3: Which decision support model should be used to address the gap?

Furthermore, this study was conducted based on a literature review. It uses Kitchenham et al. methodological guidelines [40] [41]. The relevant literature on the Fit-Gap Analysis stage, ERP customisation, and ERP misfit were reviewed, summarized and supplemented with information available from case studies and surveys. We also identified the recommendations to be considered before the Fit-Gap Analysis stage and the most important parameters to bridge gaps through the literature. We could establish this study by following three main steps: planning, conducting, and reporting.

Moreover, theoretical findings related to Fit-Gap Analysis were combined with ERP practical recommendations to derive insights.

Subsquently, we conducted interviews with experts in ERP systems: consultant in the digital transformation of financial services, legal services, and HR management, SAP BO Project Manager, and Associate Manager SAP – S/4HANA certified. The interviews were based on a questionnaire (Appendix).

Fig. 3 explains the methodology followed in diagram form.

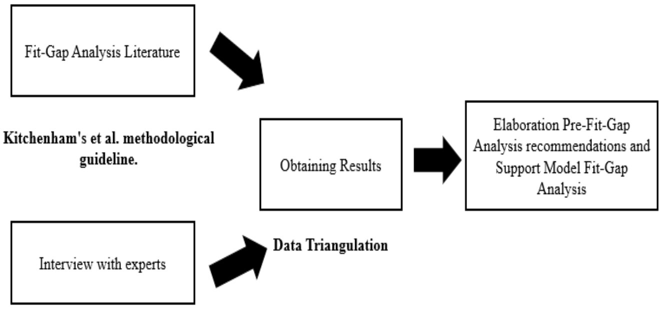

Fit-Gap Analysis Literature
Kitchenham's et al. methodological guideline.
Obtaining Results
Elaboration Pre-Fit-Gap Analysis recommendations and Support Model Fit-Gap Analysis
Interview with experts
Data Triangulation

Fig. 3. Illustration of the Methodology Followed.

To establish a model for companies and consultants, some key indicators need to be defined:

- accessible and understandable,
- easy to understand and implement, and
- representative of reality and adaptable to implementation projects.

IV. Result and Discussion

After the literature and experts recommendations, Pre-Fit-Gap Analysis recommendations are established to consult and consider before initiating the Fit-Gap Analysis. An approach and a decision support model to support stakeholders during the Fit-Gap Analysis are conceptualised using BPMN. Fig. 4 presents the study results.

This study will support the consultants and the customers throughout the Fit-Gap Analysis phase consequently the failure rate of the implementation project will be reduced. The first step is to give them recommendations to consider before starting the Fit-Gap Analysis stage to avoid redundant mistakes in previous implementation projects. The second step is to present an approach and a decision support model to prepare them for the Fit-Gap Analysis phase to bridge gaps.

A. *Pre-Fit-Gap Analysis Recommendations*

Pre-Fit-Gap Analysis recommendations are intended for both the customer and the consultants. They will be handy to prepare the stakeholders for the Fit-Gap Analysis phase and support them in making the most rational decisions, thus, avoiding mistakes that could lead to failure. Pre-Fit-Gap Analysis recommendations are presented and discussed in the eleven recommendations below.

1) Minimise customisation while trying to realise that there is a trade-off between effort and value: As mentioned, several times in the literature, customisation generates an additional cost and increases the implementation project duration. Studies have converged on a single result: customisation should be minimised to succeed [42] [43] [3]. T. Sommers and K. Nelson analysed the responses of 86 organisations, confirming this result [3].

S. Parthasarathy and S. Sharma have shown through their studies that low customisation would increase ERP efficiency. According to this study, the best situation for better system efficiency in terms of productivity is adopting the ERP standards, with minimal customisation [37].

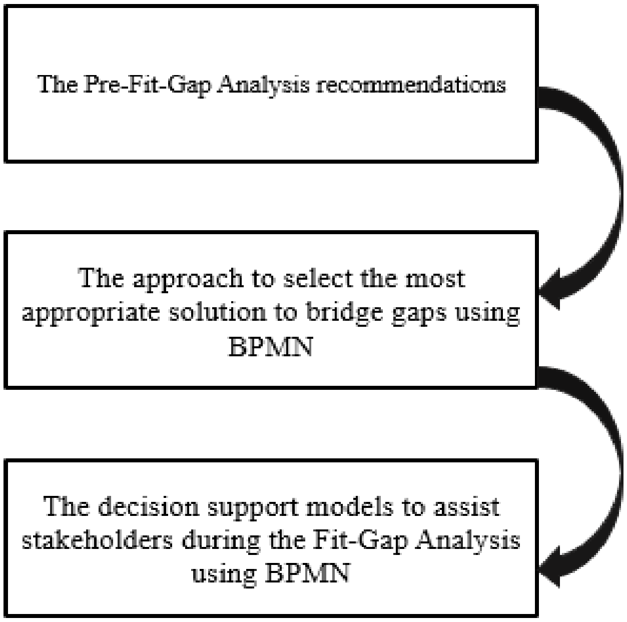

The Pre-Fit-Gap Analysis recommendations
The approach to select the most appropriate solution to bridge gaps using BPMN
The decision support models to assist stakeholders during the Fit-Gap Analysis using BPMN

Fig. 4. Illustration of the Study Results.

Customisation risks are that it introduces errors into the system, making updates more difficult and increasing maintenance costs [2] [33]. A study result made by S. Koch and K. Mitteregger shows that there is a correlation between customisation and ERP support effort. Therefore, increasing the customisation degree will increase the support effort specifically for the help desk [33].

It is clear from the previous studies regarding ERP customisation and the success factors of the implementation project that minimal customisation would be the most appropriate for ERP implementation success. Therefore, before starting the Fit-Gap Analysis phase, both the business team and the consultants should be aware that customisation should be minimised. It should be used only in case of extreme necessity and in exceptional circumstances, such as losing competitiveness in the market [37].

It is, therefore, making a compromise between effort and benefit by trying to minimise the ratio between effort and benefit. However, the effort has been defined in several ways and estimated with different methods. S. Koch and J. Mitlöhner proposed a study reviewing the literature on effort estimation methods for ERP projects, their validations, and limitations [32].

2) Train the internal team on the operation and ERP systems functionalities: Several studies have investigated the Critical Success Factors (CSF) of the ERP implementation project [3] [4] [5] [6]. Given the number of failures that have occurred over the years, Muscatello et al. reported that 40% of the implementation projects are partially successful while 20% fail [44].

Among the CFSs found in the literature that leads implementation projects to success, there is training [45]. This is a vital component of the project. However, there are two stages of training. A training session for the business team before starting the Fit-Gap Analysis phase, and another training session for end-users. Both steps are crucial for the implementation project to be successful. However, this article is more concerned with the first one.

Training can also increase user satisfaction. It is an investment that brings individual benefits and advantages to the entire company [5]. Indeed, this concerns SMEs and large organisations. Laukkanen et al. reported in their study that SMEs have poor knowledge of the ERP systems and that this should change [46].

The Business team should be trained on the ERP systems and their functionalities before starting the Fit-Gap Analysis phase to have a positive attitude towards its implementation. This would increase its acceptance by project stakeholders and end-users [47]. And therefore, training is an excellent approach to reducing resistance to change [46], thus avoiding over-customisation. It is also a way of involving the users and the business team in the implementation project, especially the Fit-Gap Analysis phase, that leads to better system optimisation and consequently a better use of the system [48].

3) Ensure operational departments' involvement: Department managers should be involved in the implementation project and especially Fit-Gap Analysis. It is not the consultants who have to decide on ERP customisation. The internal experts of the company should also be involved in this decision. The idea is not to rely heavily on the consultants, but to lead the project together by engaging the internal team more [43].

The operational department's involvement would conduct to a better ERP acceptance. This would lead to project ownership and participation in the organisation's improvement and subsequently reduce resistance to change. The company will ensure that the project is both technical and managerial success.

4) Work in harmony between internal and external teams: A Fit-Gap Analysis is carried out in teams by several stakeholders. To succeed in the Fit-Gap Analysis and achieve the fixed goals, it is essential to have cohesion within the team.

It is challenging to build the client's trust in external experts especially when client believes that they will maximise customisation to increase implementation cost since it is billed by the hour [43]. However, trusting the consultants is essential because they know the ERP systems better and can suggest improvements to the business.

Trust can only be established between internal and external teams by following certain aspects:

- having greater transparency of both parties, and
- making sure that the consultants have enough experience and knowledge of the business.by putting in place business team training.

Conflict management is also an important factor in Fit-Gap Analysis success. It allows for better group cohesion and better trust between stakeholders which opens the possibility for several solutions to the problems [49].

G. Chen et al. considered three approaches to conflict management: cooperative, competitive, and avoiding [50]. The cooperative approach allows better teamwork based on mutual help and spontaneous communication between team members. This approach also permits an exchange of ideas that stimulates creativity and innovative solutions [49].

The competitive approach is another alternative used when leaders expect very high performance. They prefer to use conflict to choose the best-performing ideas. This approach creates a competition between the stakeholders which limits the Fit-Gap Analysis phase's success [49].

The avoiding approach is based on solving problems early before they become serious issues. Leaders who use this approach tend to pay special attention to the professional needs of their subordinates. This is a conflict prevention approach. However, this method does not resolve the conflict but dissipates it [49].

It is better to adopt the cooperative approach that will push for better cohesion and communication and stay away from the avoiding approach that limits stakeholder's productivity. This way, leaders will have high expectations; hence they will employ the competitive approach which can be beneficial if used moderately [49].

5) Know the technical implementation type desired by the client: After selecting the solution that will best fit the customer's needs, a selection of deployment options is then made. The company faces several deployment options that highly depend on the system's location and other criteria chosen by the client. These criteria can be IT footprint reduction, customisation degree estimated, or update frequency (SAP Consultant). It is, therefore, crucial to be aware of deployment options chosen by the company before starting the Fit-Gap Analysis phase, as it has a direct relationship with it. But before explaining this point further, we need to get acquainted with the existing deployment options by taking the example of a leader in the ERP market, which is SAP.

There are three types of deployment options that SAP offers to its customers, on-premise, hosted private cloud, and public cloud. With an on-premise deployment, the customer deploys the licensed SAP software in its on-premise data center, i.e. the ERP is loaded and runs on the company's infrastructure, such as servers, networks, computers, etc. [46]. The customer opts for perpetual user rights for the software. He also becomes responsible for the associated hardware, implementation, and ongoing operations. Besides, on-premise ERP offers freedom of customisation [46] [51].

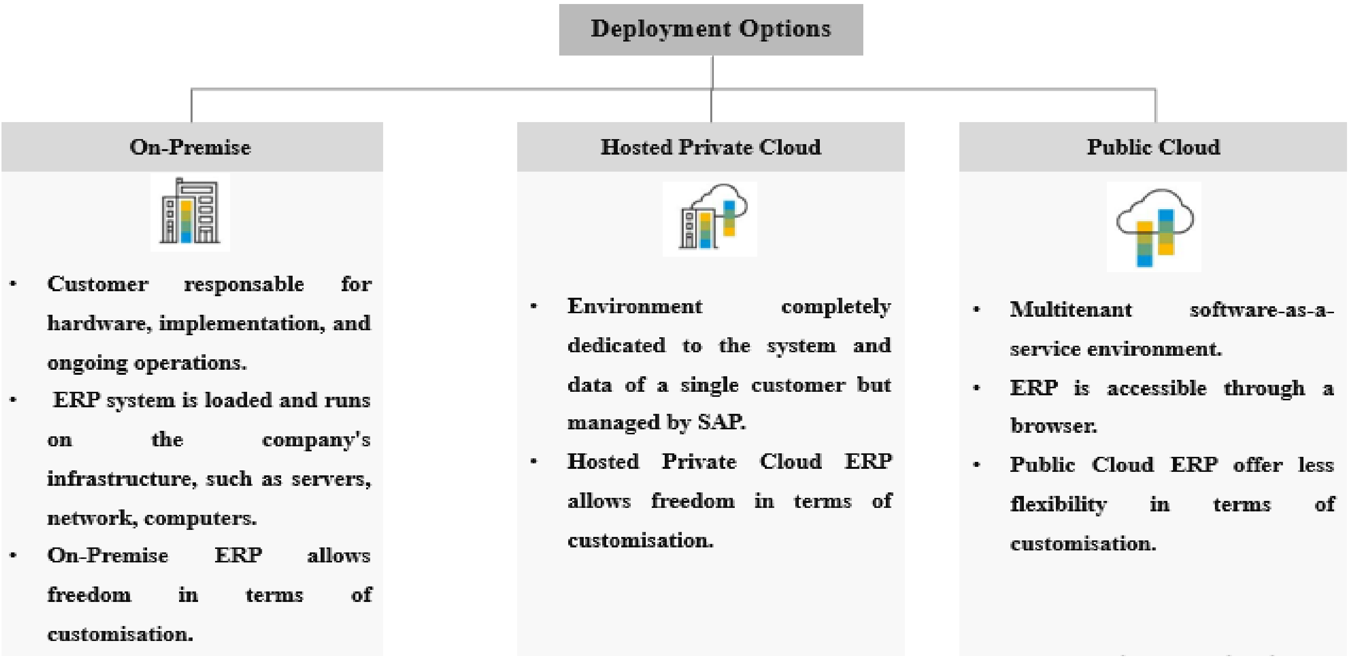
The second option is hosted private cloud, which is an environment entirely dedicated to the system and data of a single customer but managed by SAP. It is a deployment either through a perpetual license or a subscription commitment. That means the company can have its license or buy it from SAP and host it at SAP. SAP, in this case, also offers management and application service. Alternatively, the company may opt for a subscription but still decide to have its dedicated environment. In this deployment option, the company also has freedom of customisation (SAP Consultant).

The third option is the public cloud. This is a shared environment between several customers, and the software is present as Software As A Service (SAAS). SAAS is accessible through a browser [5]. Here, the company opts for a subscription-based commitment while SAP manages the software. In this deployment, the customer does not have much freedom of customisation because the source code is shared between several customers (SAP Consultant).

We have seen that some deployment options present more freedom of customisation, namely on-premise and hosted private cloud deployment, while SAAS offers less flexibility in terms of customisation. That's why it is important to know what deployment type the company has opted for before starting the Fit-Gap Analysis phase. There is no point in considering source code customisation if it is not possible with a SAAS deployment. Fig. 5 explains the different deployment options offered by SAP.

6) Evaluate the time and budget required for the implementation project: A project is a set of actions to be carried out in a predetermined time, putting in place human and material resources, that are budgeted for, and resulting in a pack of deliverables [2].

In any project, it is a question of time and budget. Before starting the Fit-Gap Analysis phase, it is essential to know the pre-defined time and budget to not exceed them. It is a question of ensuring that the solutions were chosen in the Fit-Gap Analysis phase regarding these two criteria.



- Customer responsable for hardware, implementation, and ongoing operations.
- ERP system is loaded and runs on the company's infrastructure, such as servers, network, computers.
- On-Premise ERP allows freedom in terms of customisation.

- Environment completely dedicated to the system and data of a single customer but managed by SAP.
- Hosted Private Cloud ERP allows freedom in terms of customisation.

- Multitenant software-as-a-service environment.
- ERP is accessible through a browser.
- Public Cloud ERP offer less flexibility in terms of customisation.

Fig. 5. Deployment Options.

Customisation may be seen by some as a solution that does not require an enormous cost in the short term. But it is a tremendous investment in the long- term compared to a BPR. If the company cannot implement BPR, it commits to allocating resources in the long term after the ERP is implemented [52]. To control the Total Cost of Ownership (TCO), the company should follow a structured framework to justify the radical changes that the ERP imposes [52].

Before considering any solution to bridge the gap, it is required to evaluate its cost and duration. It is, therefore, necessary to know the time and cost in the short and long term that each solution will generate to respect and best utilise the client's resources.

7) Learn about the standards proposed by the vendor: ERP systems are integrated software composed of business processes that are recognised as best practices in the industry [44]. It is critical to be aware of these standards that the ERP offers to aim at improving the company's business.

Indeed, these standards can improve business processes because they are the best practices identified across many industries. It is not only a matter of avoiding costly and risky customisation, but about improving the company's performance by adopting these practices. For this reason, it is decisive to choose the ERP that contains the standards that best fit the company's business. This is done by implementing an ERP systems selection strategy. For instance, ERP selection criteria using critical decisions analysis [53].

Before starting the Fit-Gap Analysis phase, it is essential to determine all possibilities offered by the chosen system, especially if it improves the company's performance.

8) Know flexibility degree of the system: Flexibility has been a multidimensional construct, representing the ability of a system to make the necessary readjustments to respond to environmental changes without making significant sacrifices to the company's performance [54]. Then, it is simply up to the adaptability of the system to change.

M. W. Mudie and D. J. Schafer studied ERP flexibility as an information system and not just its technical aspect [55]. They, therefore, cited two of its major components, the technical infrastructure, and the human infrastructure. The technical infrastructure is composed of applications, data, and technological configuration, while the human infrastructure is composed of knowledge, and skills required for the effective management of information technology resources in the organisation [56].

This article focuses on the technical aspect of flexibility as an element to be considered before the Fit-Gap Analysis phase. According to Byrd and Turner, the technical infrastructure flexibility includes integration and modularity [56].

Thus, before starting the Fit-Gap Analysis phase, it is crucial to evaluate the modularity aspect of the system. Duncan defined modularity as the ability to add, change and remove any software, hardware, or infrastructure data component with ease and without overall damage [57]. The idea is to evaluate what options the system offers in terms of configuration in case of a gap before considering source code modification. Once again, it is about studying all the possibilities available to resolve the gap before considering source code customisation.

9) Know the company's information technology infrastructure (IT infrastructure): IT infrastructure is the portfolio of resources that are used and shared by the company, whether at a technical or organisational level [58], allowing data to be exchanged within and with other companies. The company's IT infrastructure enables data development and use and the anticipation of future business requirements [55].

It is, therefore, significant to determine the IT infrastructure that the firm has to recognise the company's resources in human and technical capital before starting the Fit-Gap Analysis phase. This would affect solution selection to bridge the gap. Sometimes, the company cannot consider a BPR solution if it does not have certain human resources. This is also the case for the ERP customisation that requires some IT knowledge that the company should have or seek to have.

10)Know the Fit-Gap Analysis phase impact on the system maintenance: Before discussing the relationship between system maintenance and the Fit-Gap Analysis phase, it is important to define what system maintenance is. After the ERP systems implementation, it then moves to the next phase, the operation, and maintenance phase. According to Stefan Koch and Kurt Mitteregger, system maintenance begins with the vendor and first version delivery and ends with the entire product retirement. It comprises all changes to the system after it is operational. However, a distinction should be made between corrective, adaptive, and product care maintenance [33].

There is an unavoidable link between maintenance and ERP implementation projects. Hence there is a relationship between the Fit-Gap Analysis phase and maintenance [52]. The annual maintenance cost can reach up to 25% of the implementation project. This is partly because of customisation, which causes additional long-term maintenance costs. ERP vendors neither encourage this practice nor support any customisation requested by the customer [29]. The maintenance is costly because of customisation and is entirely borne by the customer [52].

It is essential to know that with each update or improvement applied to the ERP systems, all customisation should be reviewed, reapplied, and retested [59]. This becomes more complex when the person who built the customisation is not available to review, reapply or retest it. If the customer decides not to apply these updates, he may be held responsible for all the bugs in the system [52].

The literature has also emphasised the importance of communicating and collaborating with external stakeholders, especially the ERP system's vendor, to maintain the system successfully [33].

The customer should not forget that even with successful implementation, improper management of its maintenance can be costly, and can even prevent the customer from realising

ERP systems benefits. It can also cause daily transactions to fail [60].

Before starting the Fit-Gap Analysis phase, it is important to know the maintenance cost, the strategy to deal with it, and to consider it when choosing a solution during the Fit-Gap Analysis phase to estimate long-term TCO.

11)Know the company's operational characteristics complexity: The company's operational characteristics have been defined as the quantity and/or data complexity to be processed, the reports produced, the databases, and the operational processes interacting with other operational processes of different departments. For a multinational company, operational characteristics complexity increases. This is normal given data size and the frequent interaction with other business units, customers and supplier numbers [35].

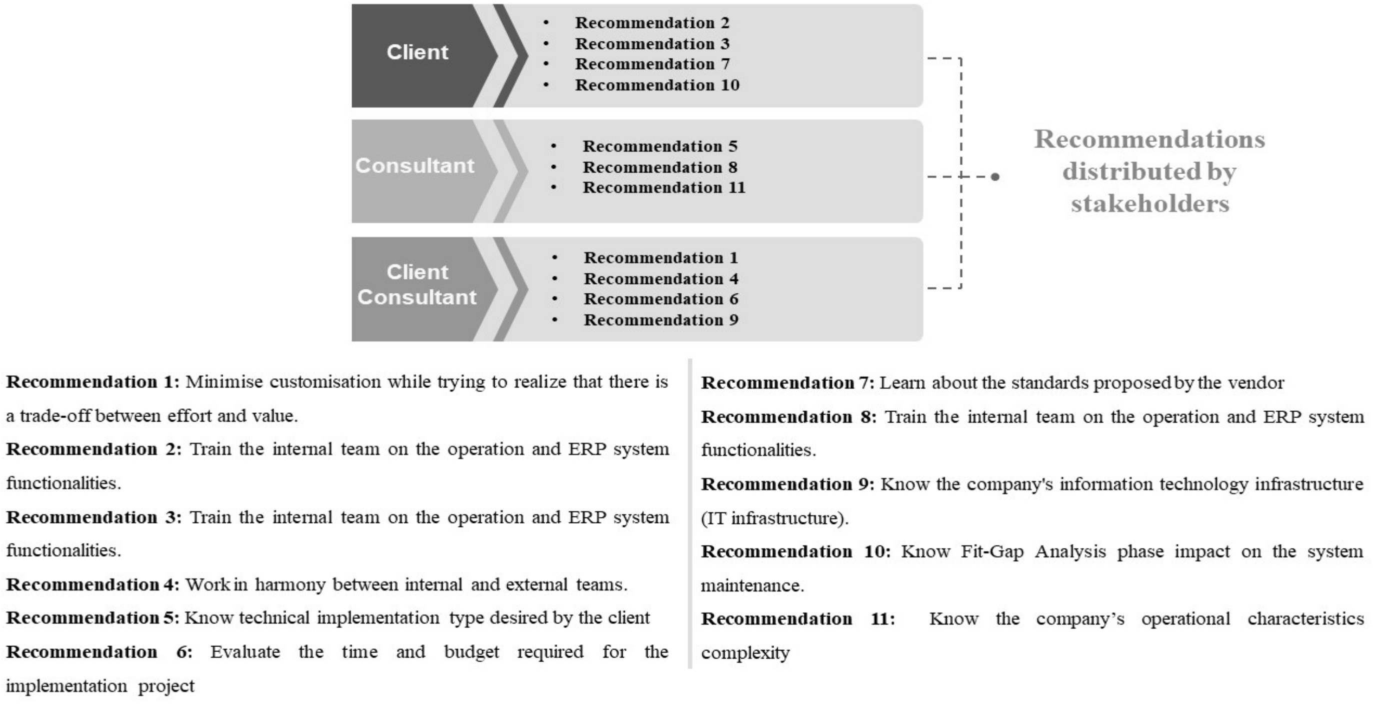
Ragwusky and Gefen have shown that ERP systems for large structures are better suited and adjusted for multinationals, hence companies with high operational characteristics' complexity [61]. This was confirmed by the study conducted by Celeste See-Pui Ng. It showed that when operational characteristics complexity increases, this is associated with a better fit with the ERP systems, whether it is in terms of processes, interfaces, or data [35]. This result should be taken into consideration by the stakeholders. Before starting the Fit-Gap Analysis phase, it would be interesting to know the company's operational characteristics complexity to estimate the fit between the ERP systems and the customer's needs.

Fig. 6 lists all the pre-Fit-Gap analysis recommendations according to the different stakeholders.

B. *Approach to be followed during the Fit-Gap Analysis Phase*

To assist stakeholders to manage the Fit-Gap Analysis phase, we designed an approach based on the findings of D. Pajk and A. Kovačič [8] and Gattiker & Goodhue [17] (Fig. 7 and Fig. 8).

First, a comparison between the customer business need and ERP systems capability to identify a set of misfits and the fits should be done. We are not going to treat the fit case because it does not pose a problem. We are going to deal with the misfit case. It is a question of categorising it into perceived misfit or actual misfit. In case it is a perceived misfit, i.e., either resistance to change, wishful thinking, or ignorance of how the ERP works, it should be solved through training and human resource management. If it is a real misfit, the proposed model should be as follows.



Recommendation 1: Minimise customisation while trying to realize that there is a trade-off between effort and value.

Recommendation 2: Train the internal team on the operation and ERP system functionalities.

Recommendation 3: Train the internal team on the operation and ERP system functionalities.

Recommendation 4: Work in harmony between internal and external teams.

Recommendation 5: Know technical implementation type desired by the client

Recommendation 6: Evaluate the time and budget required for the implementation project

Recommendation 7: Learn about the standards proposed by the vendor

Recommendation 8: Train the internal team on the operation and ERP system functionalities.

Recommendation 9: Know the company's information technology infrastructure (IT infrastructure).

Recommendation 10: Know Fit-Gap Analysis phase impact on the system maintenance.

Recommendation 11: Know the company's operational characteristics complexity

Fig. 6. Pre-Fit-Gap Analysis Recommendations according to the Stakeholders.

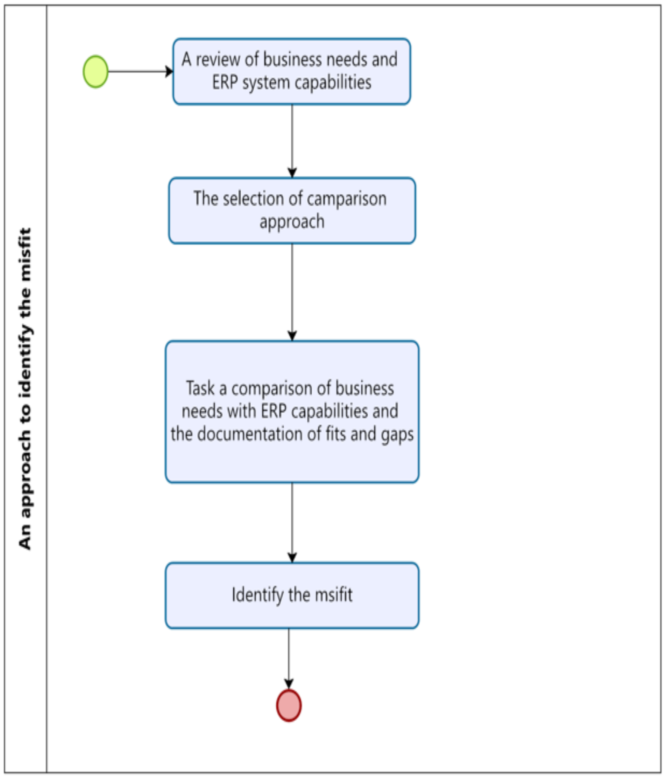

An approach to identify the misfit
A review of business needs and ERP system capabilities
The selection of camparison approach
Task a comparison of business needs with ERP capabilities and the documentation of fits and gaps
Identify the msifit

Fig. 7. An Approach to Identify the Misfit according to D. Pajk and A. Kovačič.

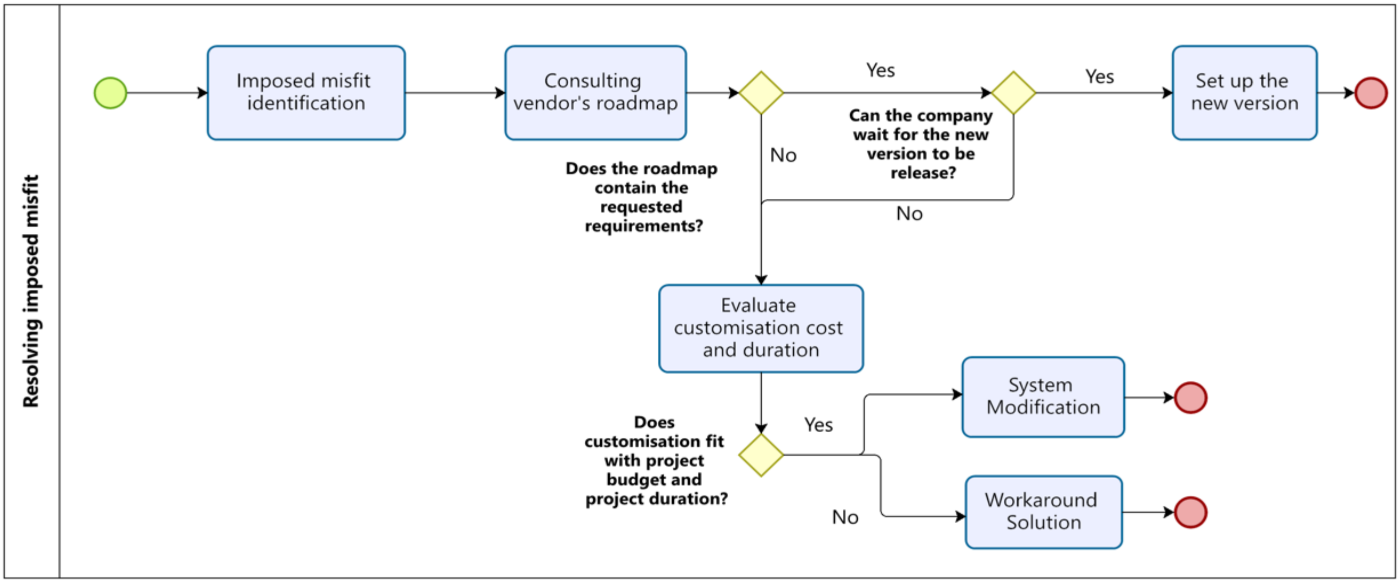

Resolving imposed misfit
Imposed misfit identification
Consulting vendor's roadmap
Does the roadmap contain the requested requirements?
Yes
No
Can the company wait for the new version to be release?
Yes
No
Set up the new version
Evaluate customisation cost and duration
Does customisation fit with project budget and project duration?
Yes
No
System Modification
Workaround Solution

Fig. 8. The Proposed Approach to Selecting the Appropriate Method to resolve the Gap.

Fig. 8 shows the approach to be followed in the Fit-Gap Analysis phase according to D. Pajk and A. Kovačič [8] and Gattiker & Goodhue [17] and subsequently proposed model application to bridge the gap.

C. Decision Support Model

We conceptualised a decision support model to support the stakeholders to manage gaps (Fig. 9, Fig. 10, Fig. 11 and, Fig. 12 in Appendix). It is based on the misfit type, the development cost, the implementation project time, the vendor’s roadmap, the uniqueness of the business processes, the criticality, frequency of the tasks, human infrastructure, and finally the data compatibility with the partners.

To apply the model, we require categorising the identified misfit. First, we need to know whether the misfit is perceived or real. Once in front of an actual misfit, it needs to be sub-categorised as either imposed misfit or voluntary misfit.

1) Imposed misfit: We are going to initially deal with the first case, which is the imposed misfit (Fig. 9 in Appendix). This misfit is what the authorities or industry standards impose. The company should comply with its requirements. However, customisation should be minimised. Therefore, before considering it, they should to look for other alternatives like consulting the vendor's roadmap, which may contain the imposed requirements.

If customisation is required, it should be analysed in terms of cost and time. It can be kept if it fits with the project duration and project budget. If not, a workaround can be proposed. It could be a third application or manual work. Fig. 9 (Appendix) explains the process to follow in the case of a realmisfit imposed.

2) Voluntary misfit: When it is a voluntary misfit, i.e., a misfit chosen by the company not imposed, the stakeholders should move on to another sub-categorisation and see if it is a surface misfit or a deep misfit (Fig. 10, Fig. 11, and Fig.12 in Appendix). If it is a deep misfit, it is then a misfit input or a misfit process.

In case it is an input misfit (Fig. 10 in Appendix), the vendor's roadmap should be consulted to see if it contains the requirements desired by the company. In case it does not, the task frequency and criticality concerned should be assessed. Assuming that the tasks are critical and frequent, customisation should be considered. Supposing not, the organisation should adapt to the system. Fig. 10 (Appendix) explains the model to be followed in the case of an input misfit.

Assuming that it is a misfit process (Fig. 11 in Appendix), then the vendor's roadmap should be consulted. If it does not contain the customer's requirements, the uniqueness of the business process in the market should be assessed. Supposing that it is a unique business process, customisation should be considered. If it is not, the company can start BPR while ensuring there is an adequate human infrastructure and no resistance to change. Fig. 11 (Appendix) explains the model to be taken in the process misfit case.

In the surface misfit case, it is, then, a question of misfit output (Fig. 12 in Appendix). The importance and coordination of the data format concerning the partners should be assessed.

In case the company's data format is already coordinated with the partners, customisation should be considered. Where the data format is not coordinated with the partners, the cost and time needed for the customisation should be discussed. And it may be possible to change the system and adapt it to the client's current format if this does not need an enormous investment. In case the company's data format is already coordinated with the partners, customisation should be considered. Fig. 12 (Appendix) explains the process to solve the gap in the misfit output case.

V. Conclusion and Future Work

The Fit-Gap Analysis is a crucial phase in the ERP implementation project. Identifying all the points leading to the success of the Fit-Gap Analysis is essential to technical and managerial success. Through the literature and business expert recommendations, we were able to collect a set of recommendations to be considered before the Fit-Gap Analysis phase. These recommendations should be consulted by both the consultants and the internal team.

These recommendations address the key points to consider before starting the Fit-Gap Analysis phase. They cover technical aspects, such as knowledge of the software's flexibility, technical implementation, system maintenance, minimisation of customisation, and the company's IT structure. Furthermore, a managerial aspect such as training, involvement of internal stakeholders, trust between consultants and internal team through improving communication, a better understanding of the standards proposed by the ERP, and a deeper insight of the company's functioning.

To assist the stakeholders in the implementation project to make the most rational decision in solving the gap based on the misfit type, we proposed a decision support model considering several parameters to bridge the gap: the development cost, the time of the implementation project, the vendor's roadmap, the business processes uniqueness, the adequate human infrastructure, the criticality and tasks frequency, and lastly data compatibility with the partners.

This model minimised customisation while respecting the company's resources. Therefore, before considering the system customisation, it is necessary to search for other ways to bridge the gap, such as consulting the vendor's roadmap, as it may contain the requirements requested by the customer. If customisation is considered, it is necessary to study its cost and time aspect before applying it. It is therefore improving the company's performance while bridging the gap.

This study can be improved by conducting a case study in the field to merge it. We can also address another aspect of the Fit-Gap Analysis phase. This is the evaluation of the chosen solution to bridge the gap through performance indicators and a post-Fit- Gap Analysis risk management study.

This study can also be refined by discussing the introduction of other technologies that can complement the ERP system and then solve ERP system issues, especially gaps. One such technology can be the blockchain. However, it is necessary to do an in-depth study of the blockchain technology since this technology is still in its early stages and requires seeing how it can complement the ERP system.

References

[1] P.-A. Millet, "Toward a model-driven, alignment-oriented ERP methodology," Computers in Industry, vol. 64, no. 4, pp. 402–411, May 2013.

[2] T. Shiang-Yen, R. Idrus, and W. P. Wong, "ERP Misfit-Reduction Strategies: A Moderated Model of System Modification and Organizational Adaptation," Journal of Global Information Management, vol. 21, no. 1, pp. 59–81, Jan. 2013.

[3] K.-K. Hong and Y.-G. Kim, "The critical success factors for ERP implementation: an organizational fit perspective," Information & Management, vol. 40, no. 1, pp. 25–40, Oct. 2002.

[4] A. Elragal, "The Impact of ERP Partnership Formation Regulations on the Failure of ERP Implementations," Procedia Technology, p. 9, 2013.

[5] T. Guimares, Y. Yoon, and Q. O'Neal, "Success factors for manufacturing expert system development," Computers & Industrial Engineering, vol. 28, no. 3, pp. 545–559, Jul. 1995.

[6] W. E. Hajj and A. Serhan, "Study on the Factors that Determine the

Success of ERP Implementation," Proceedings of the International Conference on Business Excellence, vol. 13, no. 1, pp. 298–312, May 2019.

[7] E. Hustad, M. Haddara, and B. Kalvenes, "ERP and Organizational Misfits: An ERP Customization Journey," Procedia Computer Science, vol. 100, pp. 429–439, 2016.

[8] D. Pajk and A. Kovačič, "Fit Gap Analysis – The Role of Business Process Reference Models", Economic And Business Review , vol. 15, no. 4, p. 20, 2013.

[9] A. Das and R. Narasimhan, "Process-technology fit and its implications for manufacturing performance," Journal of Operations Management, vol. 19, no. 5, pp. 521–540, Oct. 2001.

[10] E. Rabinovich, M. E. Dresner, and P. T. Evers, "Assessing the effects of operational processes and information systems on inventory performance," Journal of Operations Management, vol. 21, no. 1, pp. 63–80, Jan. 2003.

[11] D. L. Goodhue, B. D. Klein, and S. T. March, "User evaluations of IS as surrogates for objective performance," Information & Management, vol. 38, no. 2, pp. 87–101, Dec. 2000.

[12] D. L. Goodhue and R. L. Thompson, "Task-Technology Fit and Individual Performance," MIS Quarterly, vol. 19, no. 2, pp. 213–236, 1995.

[13] A. Hawari and R. Heeks, "Explaining ERP failure in a developing country: A Jordanian case study," J. Enterprise Inf. Management, vol. 23, pp. 135–160, Feb. 2010.

[14] P. Iskanius, R. Halonen, and M. Mottonen, "Experiences of ERP use in Small Enterprises.," Jan. 2009, pp. 5–10.

[15] S. Sia and C. Soh, "An assessment of package-organisation misalignment: Institutional and ontological structures," EJIS, vol. 16, pp. 568–583, Oct. 2007.

[16] J. A. A. Beijsterveld and W. J. H. Groenendaal, "Solving misfits in ERP implementations by SMEs," Information Systems Journal, p. 26, 2015.

[17] D. L. Goodhue and T. F. Gattiker, "Enterprise System Implementation and Use at Bryant Manufacturing: An Analysis of ERP Fits and Misfits," 2002, pp. 713–718.

[18] P. B. Seddon, C. Calvert, and S. Yang, "A Multi-Project Model of Key Factors Affecting Organizational Benefits from Enterprise Systems," MIS Quarterly, vol. 34, no. 2, pp. 305–328, 2010.

[19] C. Soh and S. Sia, "An Institutional Perspective on Sources of ERP Package-Organisation Misalignments," The Journal of Strategic Information Systems, vol. 13, pp. 375–397, Dec. 2004.

[20] Y. Wand and W. RY, "On the ontological expressiveness of information systems analysis and design grammars," Information Systems Journal, vol. 3, pp. 217–237, Jun. 2008.

[21] C. Soh, S. S. Kien, and J. Tay-Yap, "Enterprise resource planning: cultural fits and misfits: is ERP a universal solution?," Commun. ACM, vol. 43, no. 4, pp. 47–51, Apr. 2000, doi: 10.1145/332051.332070.

[22] N. Tsyen, R. Idrus, and U. Yusof, "A Framework for classifying misfits between enterprise resource planning (ERP) systems and business strategies," Asian Academy of Management Journal, vol. 16, Jul. 2011.

[23] S. Madnick, R. Wang, Y. Lee, and H. Zhu, "Overview and Framework for Data and Information Quality Research," J. Data and Information Quality, vol. 1, Jun. 2009.

[24] J. Grabis, "Optimization of Gaps Resolution Strategy in Implementation of ERP Systems:," in Proceedings of the 21st International Conference on Enterprise Information Systems, Heraklion, Crete, Greece, 2019, pp. 84–92.

[25] "Microsoft Dynamics Sure Step Implementation Methodology - TechNet Articles - United States (English) - TechNet Wiki." https://social.technet.microsoft.com/wiki/contents/articles/5750.microsoft-dynamics-sure-step-implementation-methodology.aspx (accessed Aug. 17, 2021).

[26] S. Sawyer, "A Market-Based Perspective on Information Systems Development," Communications of the ACM, vol. 44, Nov. 2001.

[27] G. Blick, T. Gulledge, and R. Sommer, "Defining Business Process Requirements for Large-Scale Public Sector ERP Implementations: A Case Study.," Jan. 2000, pp. 1203–1209.

[28] B. Light, "The maintenance implications of the customization of ERP software," J. Softw. Maint. Evol.: Res. Pract., vol. 13, no. 6, pp. 415–429, Nov. 2001.

[29] L. Brehm, A. Heinzl, and M. L. Markus, "Tailoring ERP systems: a spectrum of choices and their implications," in Proceedings of the 34th Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 2001, p. 9.

[30] M. Keil and A. Tiwana, "Relative importance of evaluation criteria for enterprise systems: a conjoint study," Information Systems Journal, vol. 16, no. 3, pp. 237–262, Jul. 2006.

[31] S. Parthasarathy, "Impact of customization over software quality in ERP projects: an empirical study," Software Qual J, p. 18.

[32] S. Koch and J. Mitlöhner, "Effort estimation for enterprise resource planning implementation projects using social choice - a comparative study," Enterprise Information Systems, vol. 4, pp. 265–281, Aug. 2010.

[33] S. Koch and K. Mitteregger, "Linking customisation of ERP systems to support effort: an empirical study," Enterprise Information Systems, vol. 10, no. 1, pp. 81–107, Jan. 2016.

[34] T. F. Gattiker and D. L. Goodhue, "What Happens after ERP Implementation: Understanding the Impact of Interdependence and Differentiation on Plant-Level Outcomes," MIS Quarterly, vol. 29, no. 3, pp. 559–585, 2005.

[35] C. S.-P. Ng, "A Case Study on the Impact of Customization, Fitness, and Operational Characteristics on Enterprise-Wide System Success, User Satisfaction, and System Use:," Journal of Global Information Management, vol. 21, no. 1, pp. 19–41, Jan. 2013.

[36] S. Parthasarathy, C. Sridharan, T. Chandrakumar, and S. Sridevi, "Quality Assessment of Standard and Customized COTS Products," Int. J. Inf. Technol. Proj. Manag., vol. 11, no. 3, pp. 1–13, Sep. 2020.

[37] S. Parthasarathy and S. Sharma, "Efficiency analysis of ERP packages—A customization perspective," Computers in Industry, vol. 82, pp. 19–27, Oct. 2016.

[38] B.-K. Yoo and S.-H. Kim, "Analysis of Impact on ERP Customization Module Using CSR Data," J. Inf. Process. Syst., vol. 17, no. 3, pp. 473–488, Jun. 2021.

[39] E. Uppstrom, C.-M. Lonn, M. Hoffsten, and J. Thorstrom, "New Implications for Customization of ERP Systems," in 2015 48th Hawaii International Conference on System Sciences, HI, USA, Jan. 2015, pp. 4220–4229. doi: 10.1109/HICSS.2015.505.

[40] Kitchenham B, Charters S. Guidelines for performing systematic literature reviews in software engineering. Keele University and Durham University; 2007.

[41] B. A. Kitchenham et al., "Refining the systematic literature review process—two participant-observer case studies," Empir Software Eng, vol. 15, no. 6, pp. 618–653, Dec. 2010, doi: 10.1007/s10664-010-9134-8.

[42] T. M. Somers and K. Nelson, "The impact of critical success factors across the stages of enterprise resource planning implementations," in Proceedings of the 34th Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 2001, p. 10.

[43] V. Botta-Genoulaz and P.-A. Millet, "An investigation into the use of ERP systems in the service sector," International Journal of Production Economics, vol. 99, no. 1, pp. 202–221, Jan. 2006.

[44] J. R. Muscatello, M. H. Small, and I. J. Chen, "Implementing enterprise resource planning (ERP) systems in small and midsize manufacturing firms," Int Jrnl of Op & Prod Mnagemnt, vol. 23, no. 8, pp. 850–871, Aug. 2003.

[45] S. C. L. Koh, A. Gunasekaran, and J. R. Cooper, "The demand for training and consultancy investment in SME-specific ERP systems implementation and operation," International Journal of Production Economics, vol. 122, no. 1, pp. 241–254, Nov. 2009.

[46] S. Laukkanen, S. Sarpola, and P. Hallikainen, "Enterprise size matters: objectives and constraints of ERP adoption," Journal of Ent Info Management, vol. 20, no. 3, pp. 319–334, Apr. 2007.

[47] D. H. Choi, J. Kim, and S. H. Kim, "ERP training with a web-based electronic learning system: The flow theory perspective," International Journal of Human-Computer Studies, vol. 65, no. 3, pp. 223–243, Mar. 2007.

[48] M. C. Boudreau, “Learning to use ERP technology: a causal model,” in 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the, Big Island, HI, USA, 2003, p. 10 pp.

[49] M. Attaran, “Exploring the relationship between information technology and business process reengineering,” Information & Management, vol. 41, pp. 585–596, May 2004.

[50] G. Chen, C. Liu, and D. Tjosvold, “Conflict Management for Effective Top Management Teams and Innovation in China*,” J Management Studies, vol. 42, no. 2, pp. 277–300, Mar. 2005, doi: 10.1111/j.1467-6486.2005.00497.x.

[51] J. Grabis, “On-Premise or Cloud Enterprise Application Deployment: Fit-Gap Perspective,” in Enterprise Information Systems (iceis 2019), Cham, 2020, vol. 378, pp. 406–423.

[52] M. Fryling, “Estimating the impact of enterprise resource planning project management decisions on post-implementation maintenance costs: a case study using simulation modelling,” Enterprise Information Systems, vol. 4, pp. 391–421, Nov. 2010.

[53] M. Noureddine and K. Oualid, “Extraction of ERP Selection Criteria using Critical Decisions Analysis,” ijacsa, vol. 9, no. 4, 2018.

[54] D. D’Souza and F. Williams, “Toward a taxonomy of manufacturing flexibility dimensions,” Journal of Operations Management, vol. 18, pp. 577–593, Aug. 2000.

[55] M. W. Mudie and D. J. Schafer, “An information technology architecture for change,” IBM Syst. J., vol. 24, no. 3, pp. 307–315, 1985.

[56] “Measuring the Flexibility of Information Technology Infrastructure: Exploratory Analysis of a Construct,” Journal of Management Information Systems, vol. 17, no. 1, pp. 167–208, Jun. 2000.

[57] N. B. Duncan, “Capturing Flexibility of Information Technology Infrastructure: A Study of Resource Characteristics and Their Measure,” Journal of Management Information Systems, vol. 12, no. 2, pp. 37–57, Sep. 1995.

[58] M. Broadbent, P. Weill, T. Brien, and B.-S. Neo, “Firm Context and Patterns of IT Infrastructure Capability (Best Paper Award),” p. 22.

[59] I. V. Yakovlev, “An ERP Implementation and Business Process Reengineering at a Small University,” Educause Quarterly, vol. 25, no. 2, pp. 52–57, 2002.

[60] S.-P. Ng and G. Gable, “Maintaining ERP packaged software - A revelatory case study,” Journal of Information Technology, vol. 25, no. 1, pp. 65–90, 2010.

[61] A. Ragowsky and D. Gefen, “What Makes the Competitive Contribution of ERP Strategic,” DATA BASE, vol. 39, pp. 33–49, Apr. 2008.

[62] S. Parthasarathy and M. Daneva, “An approach to estimation of degree of customization for ERP projects using prioritized requirements,” Journal of Systems and Software, vol. 117, pp. 471–487, Jul. 2016.

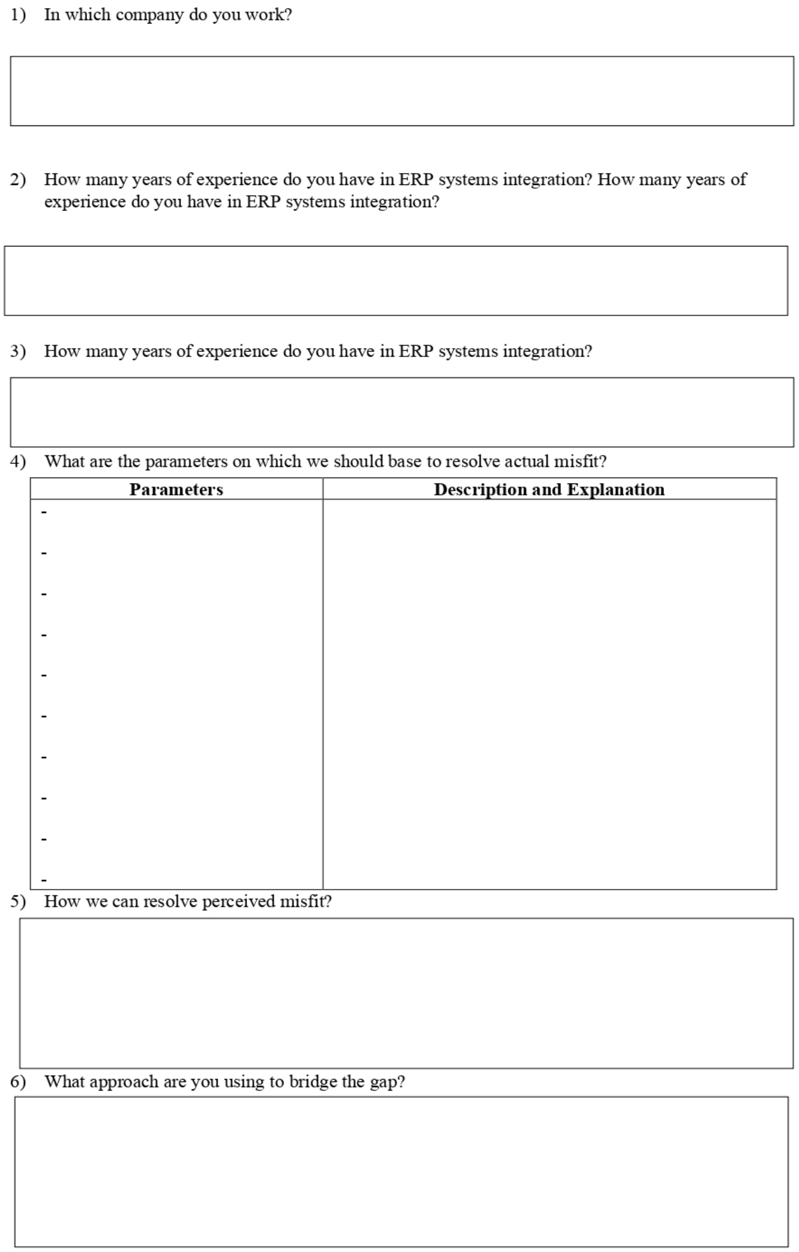
[63] S. Parthasarathy and S. Sharma, “Determining ERP customization choices using nominal group technique and analytical hierarchy process,” Computers in Industry, vol. 65, no. 6, pp. 1009–1017, Aug. 2014.

APPENDIX

TABLE I. CENSUS OF ARTICLES

Articles	Research Questions addressed by the articles
[35]	Does ERP system alignment influence system usage, user satisfaction, and system benefits?
	Do operational characteristics types influence the alignment of the ERP system, the degree of customisation, and system use?
	How does ERP system quality and/or service quality affect the relationship between ERP alignment types, operational characteristics, and degree of customisation in terms of system use, user satisfaction, and system benefits?
[33]	Does the degree of customisation have an influence on the maintenance and the degree of support? If so, is the influence positive or negative?
[8]	How do the reference models contribute to the Fit-Gap Analysis?
[16]	How do recognize the real misfit from those that are perceived?
	How to manage and understand the implementation of an ERP?
[31]	What impact does customisation have on ERP quality?
[37]	What is the relationship between ERP customisation and efficiency?
	How we can measure ERP system efficiency?
[62]	How can we know the degree of customization required for an early ERP implementatio
[7]	What are the different types of misfits? What are the types of tailoring?
	Is there a relationship between tailoring types and misfit types?
	What are the internal factors influencing making decisions process?
[39]	What changes need to be made to existing customisation options?
	What new ERP customisation options are available?
	Are the existing ERP customisation options viable for Cloud ERP?
[24]	What is the company's customisation impact preferences on the gap resolution strategy?
[63]	How we can establish customisation choices using nominal group technique and analytical hierarchy process?
[43]	Why do some companies adopt a very high level of customisation?
	What are the factors that drive ERP customisation?
	How do these factors lead to customisation?
[17]	What are the needs of consultants and software engineers in terms of requirements elicitation in the ERP domain?
	Can we provide a tool-based approach to requirements elicitation in the ERP domain?

	Can the tool-based approach support the daily work of consultants?
[28]	What customisation types are carried out to ERP software?
	Why do organisations undertake ERP software customisation?
	How might customisation impact future maintenance of the ERP software?
[33]	What is the relationship between the amount of customisation and the resulting support effort?
[15]	What are package-organisation misalignments?
	Why do package-organisation misalignments arise?
	When do organisations customise packages and when do they adapt to the package instead?
[43]	What are the factors that push ERP system customisation?
	How do these factors lead to ERP system customisation?
[49]	How do different transformational leadership behaviors influence the adoption of different conflict management methods and, consequently, influence the performance of ERP customisation projects?



1) In which company do you work?

2) How many years of experience do you have in ERP systems integration? How many years of experience do you have in ERP systems integration?

3) How many years of experience do you have in ERP systems integration?

4) What are the parameters on which we should base to resolve actual misfit?

Parameters	Description and Explanation
-	
-	
-	
-	
-	
-	
-	
-	
-	
-	

5) How we can resolve perceived misfit?

6) What approach are you using to bridge the gap?

Questionnaire:

7) What are your recommendations to consider before starting the Fit-Gap Analysis phase in order to be successful?

-

-

-

-

8) What are your recommendations to consider during the Fit-Gap Analysis phase in order to be successful?

-

-

-

-

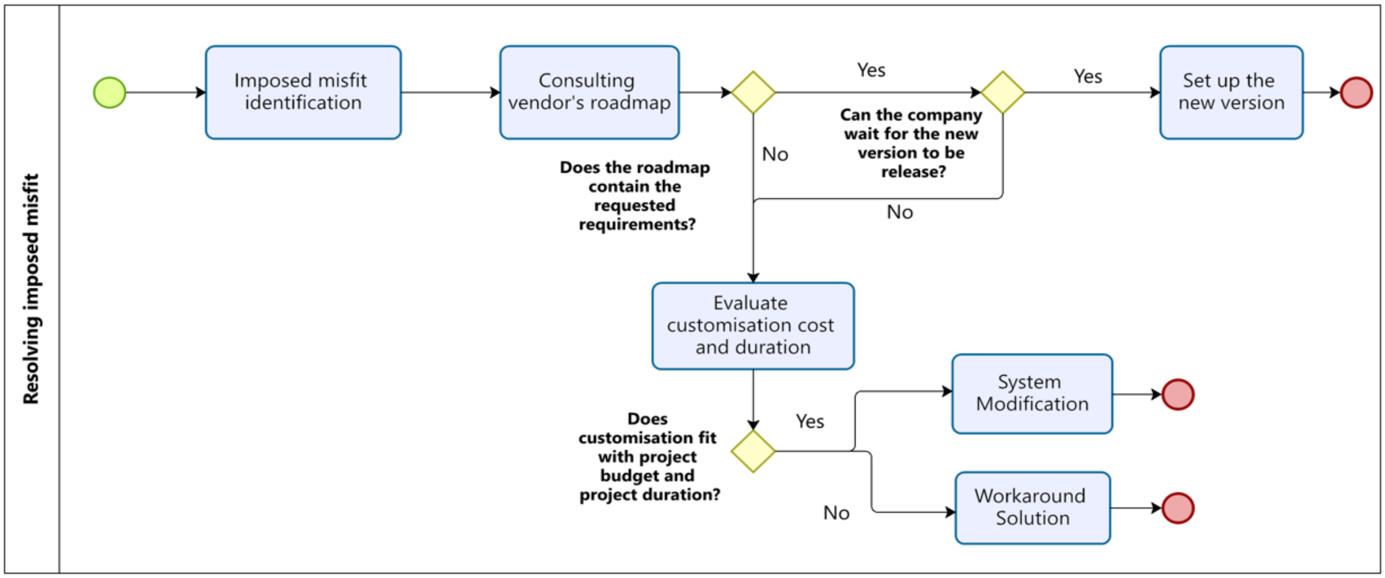

Resolving imposed misfit
Imposed misfit identification
Consulting vendor's roadmap
Does the roadmap contain the requested requirements?
Yes
No
Can the company wait for the new version to be release?
Yes
No
Set up the new version
Evaluate customisation cost and duration
Does customisation fit with project budget and project duration?
Yes
No
System Modification
Workaround Solution

Fig. 9. Resolving Imposed Misfit.

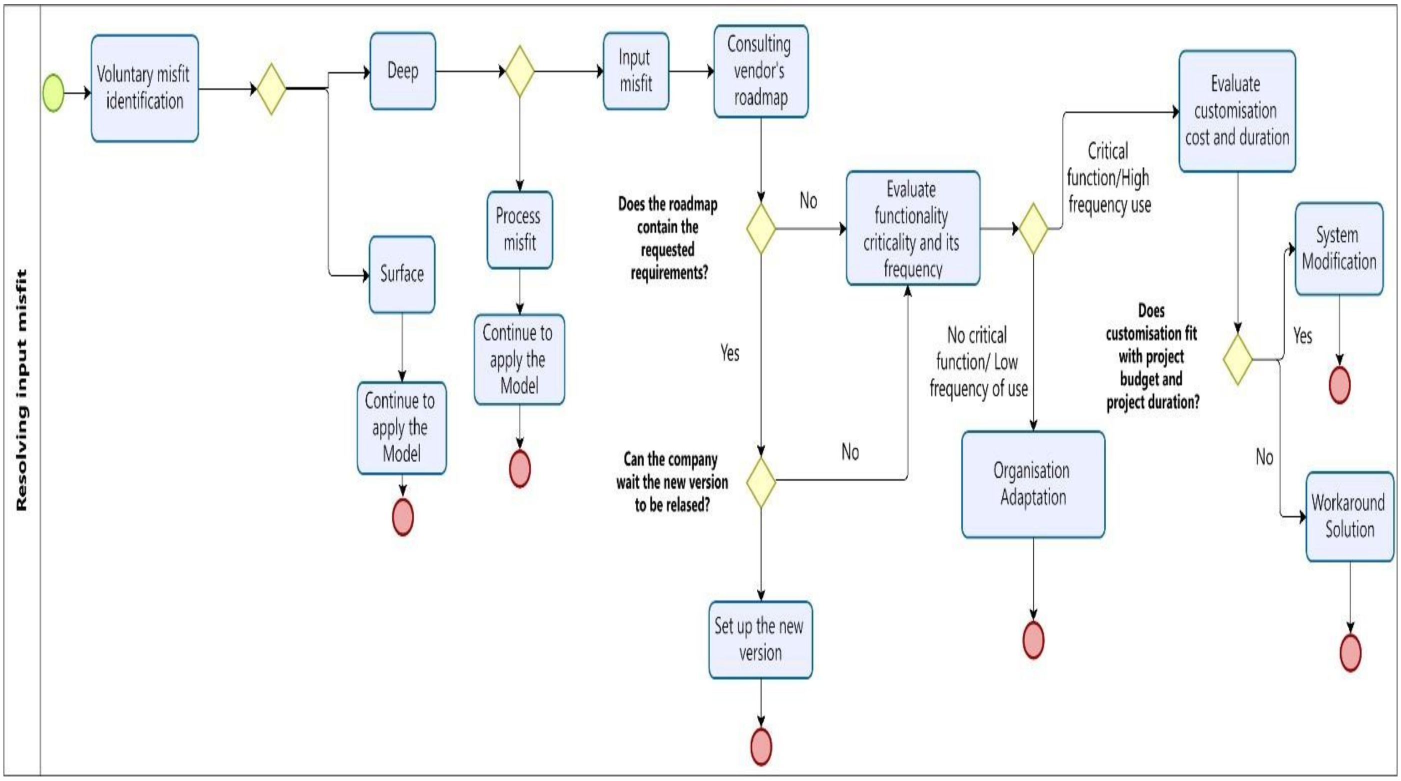

Resolving input misfit
Voluntary misfit identification
Deep
Surface
Continue to apply the Model
Process misfit
Continue to apply the Model
Input misfit
Consulting vendor's roadmap
Does the roadmap contain the requested requirements?
No
Yes
Evaluate functionality criticality and its frequency
Can the company wait the new version to be relased?
No
Set up the new version
Critical function/High frequency use
No critical function/ Low frequency of use
Organisation Adaptation
Evaluate customisation cost and duration
Does customisation fit with project budget and project duration?
Yes
No
System Modification
Workaround Solution

Fig. 10. Resolving Voluntary Misfit (Input Misfit).

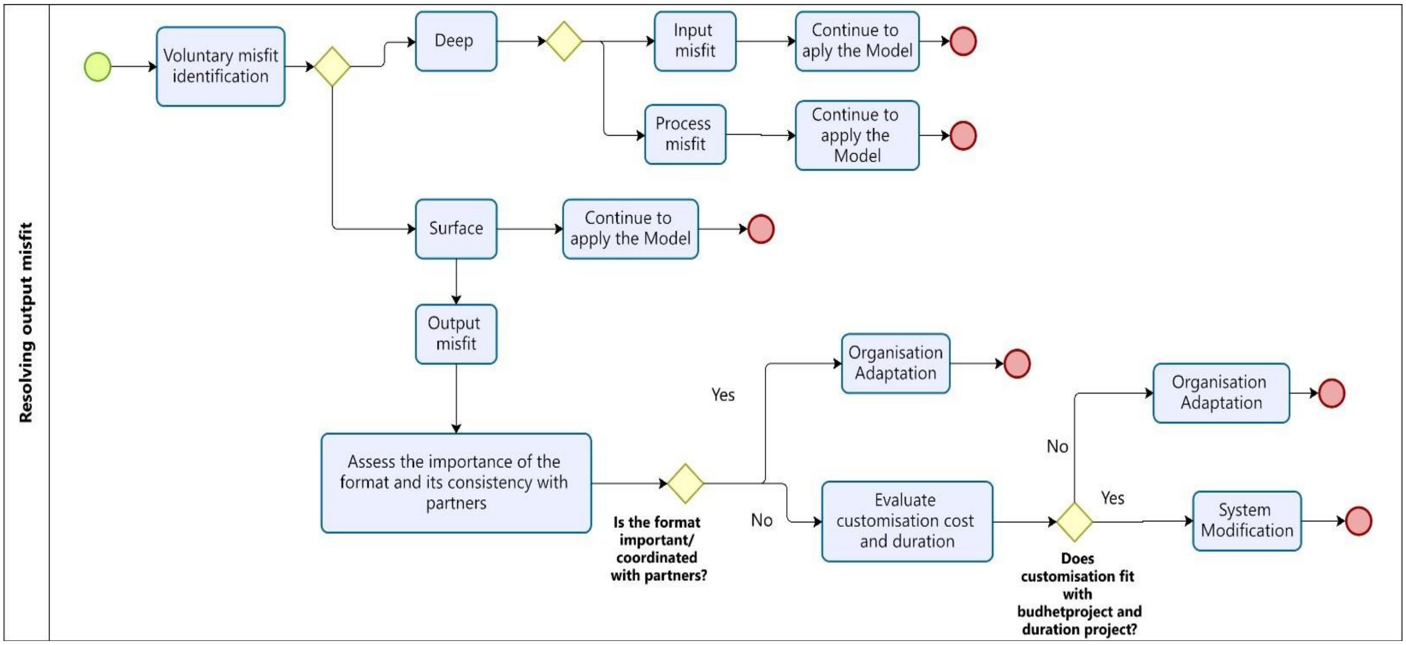

Resolving output misfit
Voluntary misfit identification
Deep
Input misfit
Continue to aply the Model
Process misfit
Continue to apply the Model
Surface
Continue to apply the Model
Output misfit
Assess the importance of the format and its consistency with partners
Is the format important/ coordinated with partners?
Yes
Organisation Adaptation
No
Evaluate customisation cost and duration
Does customisation fit with budhetproject and duration project?
No
Organisation Adaptation
Yes
System Modification

Fig. 11. Resolving Voluntary Misfit (Process Misfit).

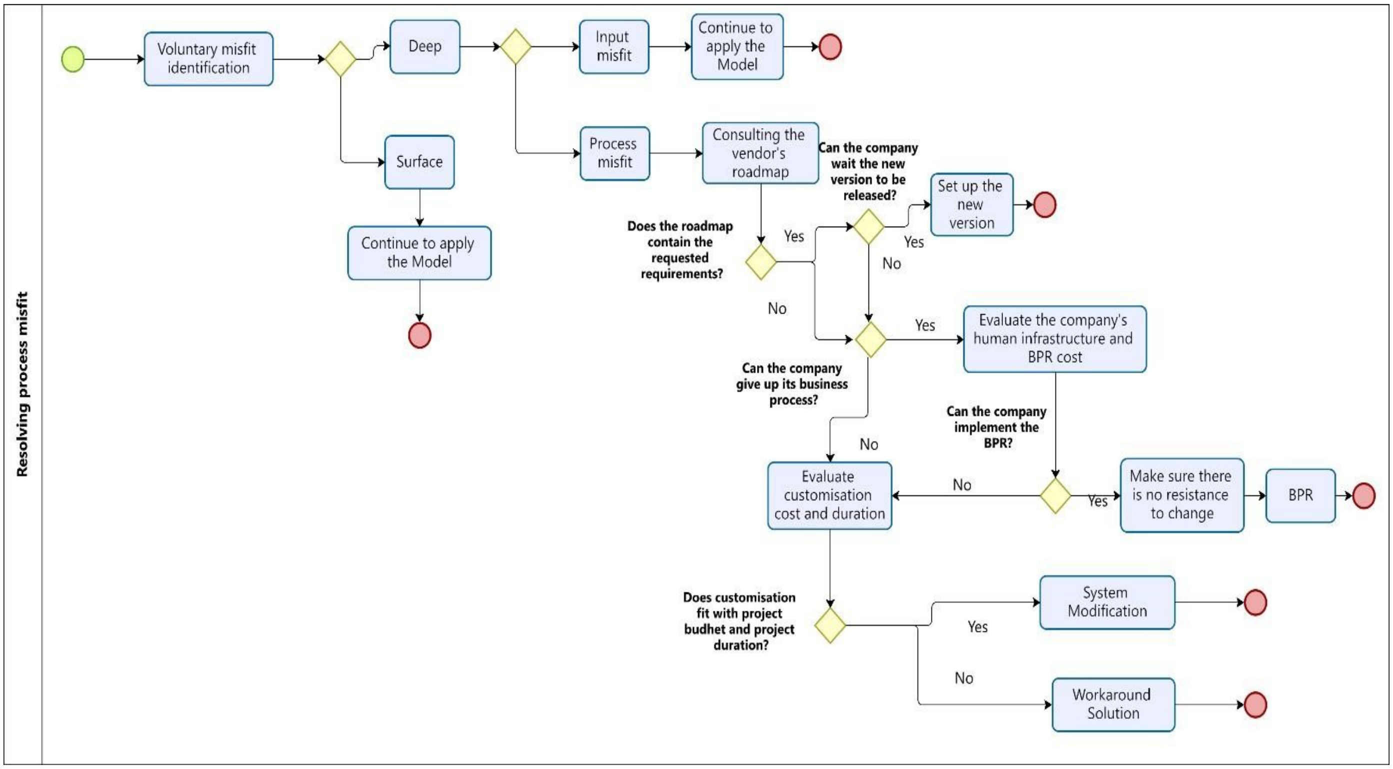

Resolving process misfit
Voluntary misfit identification
Deep
Surface
Continue to apply the Model
Input misfit
Continue to apply the Model
Process misfit
Consulting the vendor's roadmap
Does the roadmap contain the requested requirements?
Yes
No
Can the company wait the new version to be released?
Yes
No
Set up the new version
Can the company give up its business process?
Yes
No
Evaluate the company's human infrastructure and BPR cost
Can the company implement the BPR?
No
Yes
Make sure there is no resistance to change
BPR
Evaluate customisation cost and duration
Does customisation fit with project budhet and project duration?
Yes
No
System Modification
Workaround Solution

Fig. 12. Resolving Voluntary Misfit (Output Misfit).

An Improved Deep Learning Model of Chili Disease Recognition with Small Dataset

Nuramin Fitri Aminuddin, Zarina Tukiran, Ariffuddin Joret, Razali Tomari, Marlia Morsin
Department of Electrical and Electronic Engineering
Universiti Tun Hussein Onn Malaysia, Batu Pahat, 86400, Johor, Malaysia

***Abstract*—Due to its tasty and spicy fruit with nutritional qualities, chili is a demanding crop widely farmed around the world. Hence, it is essential to accurately determine the health status of chili for agricultural productivity. Recent years have seen impressive results in recognition fields due to deep learning approaches. However, deep learning models' networks need an abundant data to perform well and collecting enormous data for the networks is time-consuming and resource-intensive. A data augmentation method is proposed to overcome this problem. It was applied to a small dataset of healthy and diseased chili leaf by utilizing geometric transformation method. Eventually, two deep learning models of CNN and ResNet-18 were evaluated using augmented and original datasets. From a series of experiment, it can be concluded that the trained deep learning models using original and augmented datasets perform better with an average accuracy performance of 97%.**

Keywords—Chili leaf; deep learning; data augmentation; geometric transformation

I. Introduction

Chili (Capsicum sp.) is an important spice from the family Solanaceae that originates from South and Central America [1]. It is a demanding crop and extensively cultivated in tropical Asia and equatorial America with a high genetic diversity due to its edible and pungent fruit with nutritional values [2]. Vitamin C, potassium, phosphorus fibre, antioxidants like vitamin A, and flavonoids like β-carotene, α-carotene, lutein, zeaxanthin, and cryptoxanthin are among the nutritional values contained in a chili fruit which have the ability to suppress several human cancers [3]. However, owing to the impact of fungus, bacteria, viruses, pests, and climate on the chili cultivation process, the chili itself is susceptible to a variety of diseases. These diseases make it difficult for chili to thrive, reducing the production and quality of the fruit. It is estimated that 60-70 percent of diseases and early disease symptoms are detected just on leaves [4]. Hence, it is necessary to identify chili diseases precisely and implement early preventive and treatment measures.

Since the advent of deep learning, deep learning models have made significant advances in disease recognition [5]. Good performance from deep learning models normally needs a large number of parameters and enormous data to make these parameters operate properly. In order to do so, manual data collection and labelling [6] are required to get enormous data, which is resource-intensive and time-consuming. As a result, it can be hard to gather enough data to train the deep learning models, which significantly limits the accuracy of chili disease recognition.

With small collection of datasets, several research [7-10] in the chili agricultural field has used the data augmentation method to increase the volume of datasets. The method generates data artificially via adding augmented images to the existing dataset through either oversampling or warping [11]. By using oversampling augmentation such as generative adversarial networks (GANs), augmented images with a low likelihood of occurring label (abnormal) are added to the original datasets, preventing a deep learning model from being biassed toward the majority label of images during the recognition process. Even though GANs have intriguing promise, they need a substantial number of initial original images in order to train and create an augmented image [12]. As a result, depending on the initial size of the original dataset, GANs may not be a viable option.

In contrast, warping augmentations such as geometric transformation alter original images in such a way that their labels are maintained [13] , and this is accomplished without requiring a minimum amount of the original image to be present. Most of research [10, 13-15] that employed geometric transformation to augment original images concentrated on single transformation operation such as rotation, flipping, and scaling. To the best of the author knowledge, there has been very limited research on numerous fusions of geometric transformation operations in order to produce augmented images throughout the years. Hence, this research examines the data augmentation method known as a geometric transformation and its several transformation fusions on a small chili dataset. The augmented and original images are then fed into two deep learning models, Convolutional Neural Network (CNN) and Residual Network (ResNet-18), developed from scratch for chili disease recognition.

The contributions of this research findings can be summarized as follows. According to the findings of this research, the optimal level of accuracy for recognising chilli diseases depends on the category of datasets used and the size of the deep learning model. The finding implies that the best optimal recognition accuracy came from small datasets with both original and augmented images that were fed to a larger-sized model. This is in contrast to datasets with only original data (original datasets) and datasets with only augmented data (augmented datasets). All of the research experiments reveal that the deep learning models created from scratch are accurate to a maximum reported accuracy of 99.7%.

The remainder of this paper is organised as follows. The original dataset of healthy and diseased chili leaf produced for this research is explained in Section II. Meanwhile, Section III delves into the data augmentation method's transformation process, focusing on geometric transformation and its several transformation fusions. The architectures of deep learning models for feature extraction and recognition purposes are then discussed in Section IV. In Section V, the experimental procedures and testings used in this research to acquire the accuracy performance findings are described and the conclusion of this research is presented in Section VI.

II. Chili Leaf Image Dataset

In this research, the camera of an Oppo Reno 2 smartphone was used to capture images of chili leaf in the Batu Pahat state of Johor, Malaysia. Both types of leaf showed healthy and indications of bacterial spot disease. Only 1200 original chili leaf images were able to be acquired due to the low quantity of chilli crops in the research site. Of those 1200 images, 600 showed healthy chili leaf, while the remaining 600 showed diseased chili leaf. The images are captured in auto-focus mode at a resolution of 3000 x 4000 pixels before being resized down to 224 x 224 pixels.

III. Geometric Transformation

Data augmentation can be described as the mapping of any method that artificially increases the original dataset using the preservation label of transformations [13]:

$$\varphi = Y \rightarrow Z \qquad (1)$$

where Y is the original dataset and Z is the augmented dataset of Y. The original dataset that has been artificially increased is therefore expressed as:

$$Y' = Y \cup Z \qquad (2)$$

where Y' stores the original dataset as well as the transformations described by φ. It is worth noting that the preservation label of transformations reflects that if an image d is an element of class f, then $\varphi(d)$ is likewise an element of class f. Given that there is an infinite number of mappings $\varphi(d)$ that fulfil the criterion of preservation label of transformations, this research assesses an augmentation method, namely the geometric transformation.

Geometric transformation is a data augmentation method that alters the image's geometry by relocating the locations of each pixel's value [16]. The image's fundamental pattern of a class is preserved, but it has been shifted to a new place and alignment. This research explores the types of geometric transformation such as reflection, translation, rotation, shearing, scaling, and several fusions between them.

Reflection [17] mirrored an image around the horizontal (x-axis) or vertical (y-axis). It assists users in increasing the amount of images of an original dataset by requiring the original image matrices' rows to be inverted. In a horizontal reflection, the left and right sides of the image are turned horizontally. As shown below, the f_x and f_y components indicate the pixel's present location after reflection across the x-axis, while the coordinates of the object's original position in the image are denoted by x and y:

$$A = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \end{bmatrix} \qquad (3)$$

where A is the process equation of reflection on the x-axis. In a vertical flipping, the image is turned upside down such that the y-axis is on top and the x-axis is on the bottom. The f_x and f_y components indicate the pixel's present location after reflection across the y-axis, while the coordinates of the object's original position in the image are denoted by x and y, where B is the process equation of reflection on the y-axis:

$$B = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} . \begin{bmatrix} x \\ y \end{bmatrix} \qquad (4)$$

Then there is translation [17], which is the process of shifting an object in an image from one location to another. The translation can be performed in four directions: down, up, right and left, which helps prevent positional bias in a set of translated images. The f_x and f_y components indicate the pixel's present location after translation, while the coordinates of the object's original position in the image are denoted by x and y, where C is the process equation of translation:

$$C = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix} \qquad (5)$$

Next, rotation [18] entails spinning the original image, either in the left or right direction, with angles ranging from 1^o to 359^o. The f_x and f_y components indicate the pixel's present location after rotation while the coordinates of the object's original position in the image are denoted by x and y, where D is the process equation of rotation:

$$D = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} cos\varphi & -sin\varphi \\ sin\varphi & cos\varphi \end{bmatrix} . \begin{bmatrix} x \\ y \end{bmatrix} \qquad (6)$$

Additionally, shearing [17] is the process of altering the shape of the original image in a single direction. Shearing can be done in either the x-axis or the y-axis direction. The f_x and f_y components indicate the pixel's present location after shearing while the coordinates of the object's original position in the image are denoted by x and y.

Consequently, (7) shows the shearing in the x-axis direction, whereas (8) shows the shearing in the y-axis direction. The E and F are the process equations of images sheared on the x-axis and the y-axis directions, respectively.

$$E = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} 1 & shX \\ 0 & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \end{bmatrix} \qquad (7)$$

$$F = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ shY & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \end{bmatrix} \qquad (8)$$

In contrast, scaling [18], often known as zooming or cropping, is the process of enlarging and shrinking the original image in order to view more information. The operation of the process is to enlarge or shrink the image from a starting X, Y position to a destination X, Y. The f_x and f_y components indicate the pixel's present location after scaling, while the coordinates of the object's original position in the image are denoted by x and y,where G is the process equation of scaling.

$$G = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} Xscale & 0 \\ 0 & Yscale \end{bmatrix} . \begin{bmatrix} x \\ y \end{bmatrix} \qquad (9)$$

An object in an original image that has been reflected, whether on the x-axis or the y-axis, can be translated by shifting the reflected image into a new location, resulting in a fusion of reflection and translation transformations. Given that H and I are the process equations of images reflected on the x-axis and the y-axis, respectively, and then translated, the equations are as follows:

$$H = A \cup C \tag{10}$$

$$\mathrm{I} = B \cup C \tag{11}$$

This research also includes the fusion of reflection and scaling transformations. Given that J and K are the process equations of images scaled, and then reflected on the x-axis and the y-axis, respectively, the equations are as follows:

$$J = G \cup A \tag{12}$$

$$K = G \cup B \tag{13}$$

Additionally, this research also has a fusion of scaling and shearing transformations. Given that L and M are the process equations of images scaled, and then sheared on the direction of the x-axis and the y-axis, respectively, the equations are as follow:

$$L = G \cup E \tag{14}$$

$$M = G \cup F \tag{15}$$

Finally, there can be more than two fusions of geometric transformations. Given that N and O are the process equations of images scaled, then reflected on the x-axis and the y-axis, respectively, and lastly followed by translation, the equations are as follows:

$$N = G \cup A \cup C \tag{16}$$

$$O = G \cup B \cup C \tag{17}$$

IV. Developments of Deep Learning Model

In the leaf disease recognition domain, researchers have employed enhanced deep learning network architecture through various models [9-10] and applied them to chili disease recognition. This research employs two types of deep learning models: CNN and ResNet-18, which are developed from scratch using the Deep Network Designer [19]. An accuracy measure in [20] is used as a metric to evaluate the performance of these models. The architecture of each model is described in further detail in the following section.

A. CNN Architecture

Output, input, and hidden layers are the three primary layers of a CNN model [21]. It is common for the hidden layers to have convolutional with rectified linear unit (ReLU) function, pooling, and fully connected layers. The convolutional layer comprises a collection of filters that are used to identify features of varying sizes. Each filter convolves throughout an input image by moving horizontally for a certain amount of time, then moves vertically for another amount of time until the whole image has been convolved. A nonlinear activation function, which is the ReLU, is then applied to the convolution process' outputs. The layer which pools the neuron cluster outputs from the convolution layer into a single neuron is called the pooling layer. The pooled output is then given to a fully connected layer, which adds a bias vector and multiplies it by a weight matrix before feeding it to a softmax layer, which executes the classification operation (output). The architecture of a CNN model is shown in Fig. 1.

B. ResNet-18 Architecture

ResNet is suggested in [22] as a solution to the issues of performance deterioration and gradient vanishing caused by the depth expansion of an CNN model. Convolution layers, pooling layers, fully linked layers, softmax layers, and shortcut connections make up the architecture of a ResNet-18 model shown in Fig. 2. The shortcut connections represent the connections that travel between two layers. There are two main kinds of pooling layers in the ResNet-18 model architecture in this research. The first of these layers is the max-pooling layer, which chooses the maximum element from the area of the feature map covered by the convolution filter. For the second layer which is the average pooling layer, instead of picking the maximum element, it works by calculating the average value of the element from the region of the feature map.

The building layer of ResNet-18 is seen in Fig. 3 with an input x parameter and the desired output $H(x)$. The block makes use of a shortcut connection that enables it to immediately learn the residual $F(x) = H(x) - x$ in order to generate the desired output $[F(x) + x]$, hence avoiding performance deterioration and gradient vanishing due to an excessive number of convolutional layers.

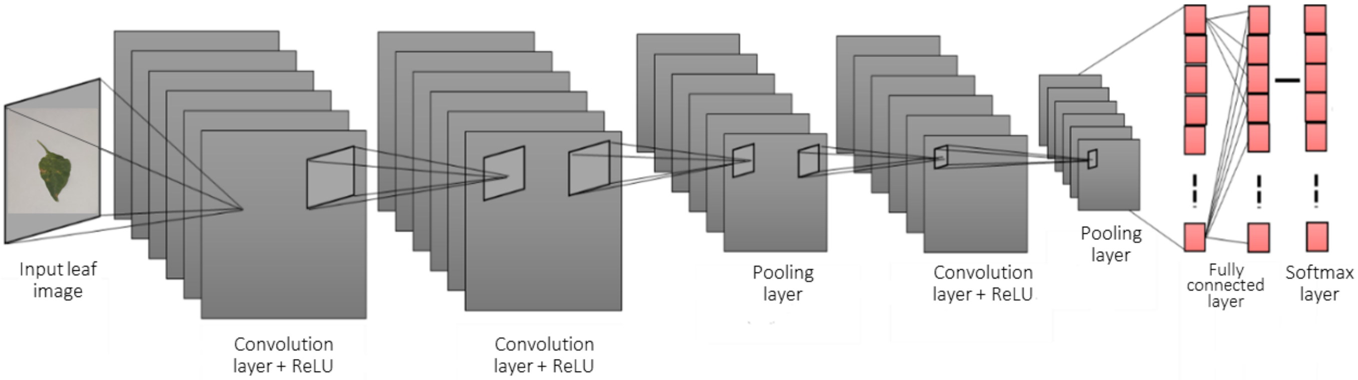


Fig. 1. CNN Model Architecture.

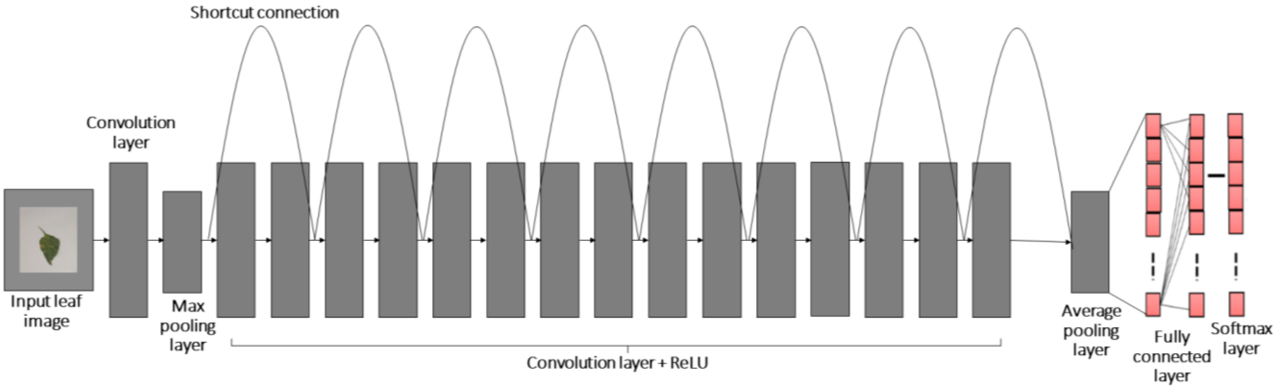


Fig. 2. ResNet-18 Model Architecture.

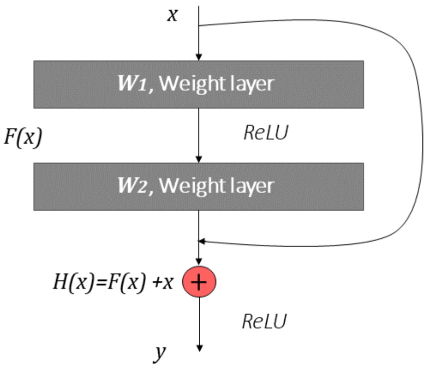


Fig. 3. The Building Layer of ResNet-18.

The ResNet building layer in Fig. 3 employs the residual mapping function [23] described below , where σ:

$$P = W_2\sigma(W_1x) \quad (18)$$

denotes the activation function of ReLU. Through a second activation function of ReLU, the output y can be obtained:

$$y = F(x, \{W_i\}) + x \quad (19)$$

A linear transformation of output y can also be obtained by multiplying W_s to x in (19) as shown below.

$$y = F(x, \{W_i\}) + W_sx \quad (20)$$

V. Results and Discussion

The accuracy performance results for both the CNN and ResNet-18 models on original and augmented datasets of healthy and diseased chili leaf are acquired via the use of experimental setup and testing, which are detailed in the following section.

A. Experimental Setup

All of the models in the experiments run on MATLAB® with an Intel® CoreTM i3 processor operating at 3.4 GHz. The models' networks are fed data from three different categories of datasets: datasets with only original data (original datasets), datasets with only augmented data (augmented datasets), and datasets with both original and augmented data (original + augmented datasets). Only 1200 images from each category datasets are fed into the models in order to preserve the data balance. Therefore, 40 images from the original datasets are chosen to be augmented, consisting of 20 random images of healthy chili leaf and 20 random images of diseased chili leaf. During the augmentation process, which comprises of 15 geometric transformations, 600 augmented images are produced and preserved in the augmented datasets, while the original images are discarded. For original + augmented datasets, 300 images of original and augmented healthy chili leaf, as well as 300 images of original and augmented diseased chili leaf, are used. Table I shows the specific information for all the datasets in this research.

TABLE I. DATASETS INFORMATION

Type of datasets	*Ori_HC	*Ori_DC	*Aug_HC	*Aug_DC	Total images
Original	600	600	0	0	1200
Augmented	0	0	600	600	1200
Original + Augmented	300	300	300	300	1200

*Ori_HC = Original images of healthy chili leaf
*Ori_DC = Original images of diseased chili leaf
*Aug_HC = Augmented images of healthy chili leaf
*Aug_DC = Augmented images of diseased chili leaf

When the CNN and ResNet-18 models are fed with a dataset, 70% of the data in the dataset is utilised to train the models, while the remaining 30% is used to test the models. During training, the hyperparameter settings [24] of both models, such as batch size, learning rate, maximum epoch, testing frequency and optimizer, are fixed such that the optimum performance of both models is equal. Table II summarises the fixed hyperparameter settings for both models.

TABLE II. FIXED HYPERPARAMETER SETTINGS

Hyperparameter setting	Value
Batch size	32
Learning rate	0.0001
Maximum epoch	30
Testing frequency	50
Optimizer	Stochastic gradient descent with momentum (sgdm)

In each experiment, the accuracy performance of a developed model given an input dataset is determined using the following formula:

$$Accuracy_{Od} = \frac{Testing\ accuracy_{Ori\_HC} + Testing\ accuracy_{Ori\_DC}}{2} \quad (21)$$

$$Accuracy_{Ad} = \frac{Testing\ accuracy_{Aug\_HC} + Testing\ accuracy_{Aug\_DC}}{2} \quad (22)$$

$$Accuracy_{OAd} = \frac{Testing\ accuracy_{Oad\_HC} + Testing\ accuracy_{Oad\_DC}}{2} \quad (23)$$

*Od = Original datasets

*Ad = Augmented datasets

*OAd = Original + augmented datasets

*Oad_HC= Ori_HC + Aug_HC

*Oad_DC = Ori_DC + Aug_DC

B. *Experimental Testing*

Fig. 4 displays the geometric transformations that were performed on images of healthy and diseased chili leaf in order to create augmented images. The performed geometric transformations are based on the process equations in section III.

The accuracy results obtained by developed models when applied to the three categories of datasets, which are referred to as original datasets, augmented datasets and original + augmented datasets are shown in Table III.

Table III shows that after applying both models to the three categories of datasets, the accuracy performance achieved for each dataset varies. The CNN model produced accuracies of 92.3%, 82.5%, and 94.2% for the original datasets, augmented datasets, and original + augmented datasets, respectively. Conversely, the ResNet-18 model achieved accuracies of 99.2%, 91.8 %, and 99.7% for the original datasets, augmented datasets, and original + augmented datasets, respectively. In every experiment undertaken, it can be concluded that the ResNet-18 model outperforms the CNN model in terms of accuracy.

Despite the fact that both models were trained on 600 images per class, the average accuracy attained from the original datasets was only 95.8%. The geometric transformation method improved performance and yielded the best accuracy result of the two models, with an average recognition accuracy of 97% from both models that can be seen from original + augmented datasets. These findings indicate that the geometric transformation method improves the abilities of the models to generalise [25] by modifying the orientation of original image while retaining its original information.

On the other hand, the accuracy results from the augmented datasets showed that if the models were only trained with augmented images, the accuracy dropped by 8.6% and 9.8% on average compared to the accuracy of the original datasets and the original + augmented datasets. This is due to the black pixel areas (background areas) in the augmented images created by geometric transformation and the absence of the attention mechanism [26] that is found in the original images. The deep learning models use more background areas of the augmented images as distinct regions in the training process, leading to lower accuracy performance.

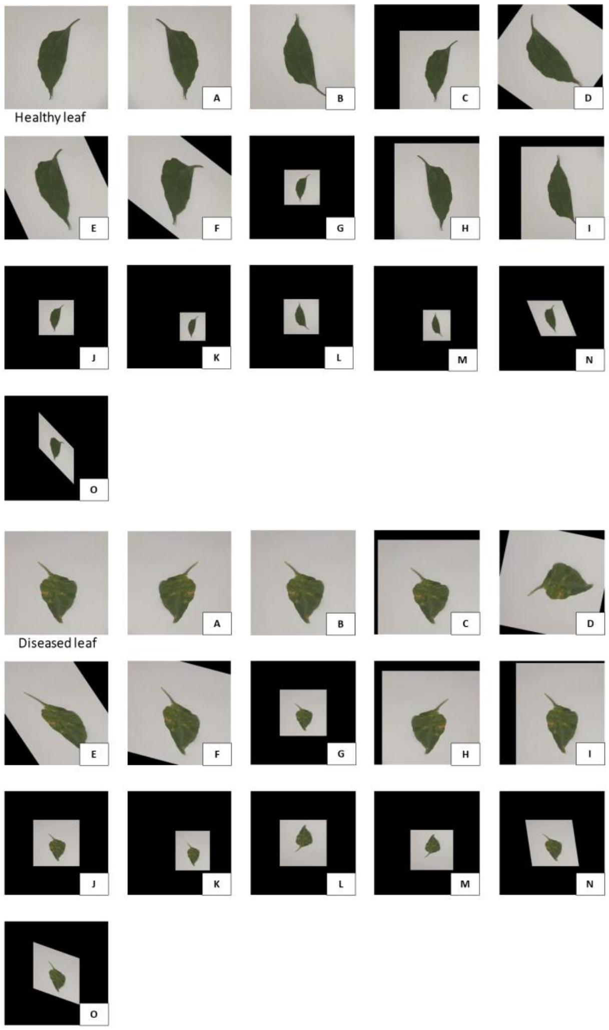


Fig. 4. Applied Geometric Transformation on Healthy and Diseased Chili Leaf Images. A) Reflection on tℎe x-axis. B) Reflection on tℎe y-axis. C) Translation. D) Rotation. E) Shearing on the x-axis. F) Shearing on the y-axis. G) Scaling. H) Reflection on tℎe x-axis and Translated. I) Reflection on tℎe y-axis and Translated. J) Scaling and Reflected on the x-axis. K) Scaling and Reflected on the y-axis. L) Scaling and Sheared on the x-axis. M) Scaling and Sheared on the y-axis. N) Scaling, Reflected on the x-axis and then Translated. O) Scaling, Reflected on the y-axis and then Translated.

TABLE III. ACCURACY RESULTS OF DATASETS

Type of datasets	Accuracy given by CNN	Accuracy given by ResNet-18
Original datasets	92.3%	99.2%
Augmented datasets	82.5%	91.8%
Original+augmented datasets	94.2%	99.7%

VI. Conclusion

This research proposed the data augmentation method known as a geometric transformation and its several transformation fusions on a small chili dataset and tested on two deep learning models, CNN and ResNet-18. A clear improvement in accuracy performance results were seen for both models after adding augmented images into the original datasets. The accuracy of both models went up by 94.2% for the CNN model and 99.7% for the ResNet-18 model.This suggests that a combination of the original and augmented images can improve the accuracy performance of the models substantially. Further research also revealed that ResNet-18 had the highest accuracy performance among both models when no data augmentation was used.

Acknowledgment

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 (vot H916). Special appreciation to the IoT Laboratory, UTHM, for providing related facilities.

References

[1] H. Thakur, S. K. Jindal, A. Sharma, and M. S. Dhaliwal, “A monogenic dominant resistance for leaf curl virus disease in chilli pepper (Capsicum annuum L.),” Crop Prot., vol. 116, pp. 115–120, 2019, doi:10.1016/j.cropro.2018.10.007.

[2] L. Colney, W. Tyagi, and M. Rai, "Morphological and molecular characterization of two distinct chilli cultivars from North-Eastern India with special reference to pungency related genes," Sci. Hortic. (Amsterdam), vol. 240, pp. 1–10, 2018, doi: 10.1016/j.scienta.2018.05.045.

[3] M. Lu, C. Chen, Y. Lan, J. Xiao, R. Li, J. Huang, Q. Huang, Y. Cao and C. T. Ho, “Capsaicin-the major bioactive ingredient of chili peppers: bio-efficacy and delivery systems,” Food Funct., vol. 11, no. 4, pp. 2848–2860, 2020, doi: 10.1039/D0FO00351D.

[4] G. Dhingra, V. Kumar, and H. D. Joshi, “Study of digital image processing techniques for leaf disease detection and classification,” Multimed. Tools Appl., vol. 77, no. 15, pp. 19951–20000, 2018, doi: 10.1007/s11042-017-5445-8.

[5] S. Mishra, A. Dash, and L. Jena, “Use of deep learning for disease detection and diagnosis,” Bio-inspir. Neurocomput. (Singapore), pp. 181–201, 2021, doi: 10.1007/978-981-15-5495-7_10.

[6] A. C. Ferreira, L. R. Silva, F. Renna, H. B. Brandl, J. P. Renault, D. R. Farine, R. Covas and C. Doutrelant, “Deep learning-based methods for individual recognition in small birds,” Methods Ecol. Evol., vol. 11, no. 9, pp. 1072–1085, 2020, doi: 10.1111/2041-210X.13436.

[7] G. Sahuri and Rosalina. “Implementation of deep learning methods in detecting disease on chili leaf,” Advanc. Stud. Comput., Sci. Eng., vol. 9, no. 6, pp. 10-15, 2020.

[8] T. L. Lin, H. Y. Chang, and K. H. Chen, “Pest and disease identification in the growth of sweet peppers using faster R-CNN,” IEEE Cons. Elect. (Taiwan), 2019, doi: 10.1109/ICCE-TW46550.2019.8991893.

[9] Y. H. Gu, H. Yin, D. Jin, J. H. Park, and S. J. Yoo, “Image-based hot pepper disease and pest diagnosis using transfer learning and fine-tuning,” Front. Plant Sci., vol. 12, 2021, doi: 10.3389/fpls.2021.724487.

[10] K. Deeba and B. Amutha, “ ResNet - deep neural network architecture for leaf disease classification,” Microprocess. Microsyst., 2020, doi: 10.1016/j.micpro.2020.103364.

[11] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” J. Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[12] S. Tim, G. Ian, Z. Wojciech, C. Vicki, R. Alec and C. Xi, “Improved techniques for training GANs,” arXiv preprint, 2016, doi:10.48550/arXiv.1606.03498.

[13] L. Taylor and G. Nitschke, “Improving deep learning with generic data augmentation,” IEEE Comput. Intel., 2018, doi: 10.1109/SSCI.2018.8628742.

[14] A. Mikołajczyk and M. Grochowski,“Data augmentation for improving deep learning in image classification problem,” Inter. PhD. Work.,pp. 117–122, 2018, doi: 10.1109/IIPHDW.2018.8388338.

[15] P. Pawara, E. Okafor, L. Schomaker, and M. Wiering, “Data augmentation for plant classification,” Inter. Conf. Advanc. Conc. Intel. Visi. Syst., pp. 615–626, 2017, doi: 10.1007/978-3-319-70353-4_52.

[16] X. Wang, K. Wang, and S. Lian, “A survey on face data augmentation for the training of deep neural networks,” Neural. Comput. Appl., vol. 32, no. 19, pp. 15503–15531, 2020, doi: 10.1007/s00521-020-04748-3.

[17] A. Vyas, S. Yu, and J. Paik, “Fundamentals of digital image processing,” Sign. Commun. Tech. (Singapore), pp. 3–11, 2018, doi: 10.1007/978-981-10-7272-7_1.

[18] L. Sifre and S. Mallat, “Rotation, scaling and deformation invariant scattering for texture discrimination,” IEEE Comput. Vision. Pattern. Recog., 2013, doi: 10.1109/CVPR.2013.163.

[19] Borda, M., Terebes, R., Malutan, R., Ilea, I., Cislariu, M., Miclea, A., and Barburiceanu, S., “Supervised deep learning classification algorithms. randomness and elements of decision theory applied to signals, pp. 205–215, 2021, doi.:10.1007/978-3-030-90314-5_15.

[20] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” Comput. Electron. Agric., vol. 147, pp. 70–90, 2018, doi:10.1016/j.compag.2018.02.016.

[21] R. B. Arif, M. A. B. Siddique, M. M. R. Khan, and M. R. Oishe, “Study and observation of the variations of accuracies for handwritten digits recognition with various hidden layers and epochs using convolutional neural network,” Elect. Eng. Inform. Comm. Tech., 2018, doi: 10.1109/CEEICT.2018.8628078.

[22] S. Kiranyaz, T. Ince, and M. Gabbouj, “Real-time patient-specific ECG classification by 1-D convolutional neural networks,” IEEE Trans. Biomed. Eng., vol. 63, no. 3, pp. 664–675, 2016, doi: 10.1109/TBME.2015.2468589.

[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Confer. Comput. Visi. Patter. Recog., 2016, doi: 10.1109/CVPR.2016.90.

[24] F. Hutter, L. Kotthoff, and J. Vanschoren, ”Automated machine learning: methods, systems, challenges,” Springer, 2019, doi: 10.1007/978-3-030-05318-5.

[25] G. Li, L. Liu, G. Huang, C. Zhu, and T. Zhao, “Understanding data augmentation in neural machine translation: two perspectives towards generalization,” Emp. Method. Nat. Lang. Process., 2019, doi: 10.1109/CVPR42600.2020.00815.

[26] H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, “Attention branch network: learning of attention mechanism for visual explanation,” Comput. Vision. Pattern. Recog., 2019, doi: 10.1109/CVPR.2019.01096.

Arabic Image Captioning: The Effect of Text Pre-processing on the Attention Weights and the BLEU-N Scores

Moaz T. Lasheen, Nahla H. Barakat
Faculty of Informatics and Computer Science, The British University in Egypt, Cairo, Egypt

Abstract—**Image captioning using deep neural networks has recently gained increasing attention, mostly for English langue, with only few studies in other languages. Good image captioning model is required to automatically generate sensible, syntactically and semantically correct captions, which in turn requires good models for both computer vision and natural language processing. The process is more challenging in case of data scarcity, and languages with complex morphological structures like the Arabic language. This was the reason why only limited number of studies have been published for Arabic image captioning, compared to those of English language. In this paper, an efficient deep learning model for Arabic image captioning has been proposed. In addition, the effect of using different text pre-processing methods on the obtained BLEU-N scores and the quality of generated images, as well as the attention mechanism behavior were investigated. Furthermore, the "THUMB" framework to assess the quality of the generated captions is used -for the first time- for Arabic captions' evaluation. As shown in the results, a BLEU-4 score of 27.12, has been achieved, which is the highest obtained results so far, for Arabic image captioning. In addition, the best THUMB scores were obtained, compared to previously published results on common images.**

Keywords—Arabic image captioning; computer vision; deep learning; image captioning; natural language processing

I. Introduction

A. Overview

Recently, automatic image captioning became a hot topic, building on the success of deep neural networks in the areas of computer vision and Natural Language Processing (NLP) tasks. Image captioning models require two main components; the first is to extract the image's features, detect its objects, and describe their relationships; while the second is the language model that converts those features to a meaningful word sequence [1, 2]. These models are initially trained on a data set of images, along with their corresponding captions [3].

Studies for caption generation is largely in English due to the availability of data sets and other pre-trained image and language models. The most commonly used architectures are Encoder-Decoder, with or without additional, optional layers, like different attention mechanisms and different embedding models [3-5]. The Encoder–Decoder architecture mainly uses several variations of a Convolutional Neural Network (CNN) as encoders, where high-level feature are extracted from the input images, which are then passed to the decoder (language) model; where Recurrent Neural Networks (RNN) have been widely used. Recently, transformers, as well as Generative Adversarial Neural Networks (GANS) models have been used. Encoder architectures included AlexNet, VGG-16 Net, RESNet, GoogleNet and DenseNet [6]. However, RESNet showed better performance, and had fewer training parameters compared to other common encoders like VGG variants [3]. For language models (the decoders), Long Short-Term Memory (LSTM), RNN, Gated Recurrent Units (GRU) have been adopted [6]. However, the LSTM is the most widely used decoder, for its ability to remember long term dependencies in the generated word sequence [6]. Several attention models have also been proposed; including hard or soft, top-down, bottom-up, semantic, and other attention methods [6, 7]. Attention methods are also used with GANs and Reinforcement Learning, which have shown excellent performance [8]. For more details on English image captioning, please refer to the reviews in [3-9]. The situation is different for Arabic image captioning, as only few models have been proposed with less satisfying results. This can be attributed to the complex morphological structure of the Arabic language, and the scarcity of data sets of images with Arabic captions. Image captioning has many valuable applications; like image indexing and retrieval, assisting visually impaired people, robot vision systems, medical image description, analysis of traffic data, and other industrial applications [10, 11].

In this paper, an efficient model for Arabic image captioning is proposed, utilizing an encoder-decoder architecture, with soft attention mechanism, and beam search to generate best captions. The paper attempts to answer the following research questions: 1) what is the effect of different Arabic text pre-processing methods on the BLEU-N scores, the behavior of the attention mechanism, and quality of the generated captions? 2) Dose beam search improve BLEU-N scores?

As the result section shows, the proposed model achieved the highest BLEU-4 score so far; for Arabic image captioning. In addition, the quality of the generated captions compares favorably to the related work, as measured by THUMB score, which is used for the first time for Arabic captions evaluations, as well as ratings of four Arabic native speakers.

The rest of the paper is organized as follows: Section B summarizes this study's contributions, followed by a review and analysis of the related work in Sections II. The experimental methodology, results and discussion, are presented in Sections III and IV respectively. Section V discusses the effect of text pre-processing methods on the attention visualization, and the paper conclusion is presented in Section VI.

B. *The Paper's Contributions*

- The proposed model in this paper for Arabic image captioning achieved the highest BLEU-4 Score so far,
- For the first time, the paper investigates the effect of Arabic text pre-processing on the attention mechanism, as well as the BLEU-N scores,
- For the first time, images' captions are qualitatively evaluated using THUMB scores,
- The paper presents the most comprehensive literature review for Arabic image captioning.

II. Related Work

The first published study on Arabic image captioning was in 2018 by [12]. Since then, the majority of the published studies used Encoder-Decoder architectures, with or without attention mechanisms; and recently, transformers have been used. The following Sections summarize the work in this area.

A. *Encoder – Decoder based Models*

The model which obtained the best results as measured by the BLEU-4 score is [12]. This model is different from all other published ones, as it uses Region Convolutional Neural Network (RCNN) to map the image objects to Arabic root words, where a transducer based algorithm has been used for this purpose. The output root words are then passed to an LSTM to generate the standard Arabic caption, and a dependency tree constraints algorithm has been used to ensure that the generated caption is grammatically correct. The authors reported the BLEU-N scores on Middle Eastern newspapers & Flickr8 data sets [12]. In [13] , a CNN was used as the encoder, and LSTM as decoder, on a part of Flickr8 data set, and used BLEU-N for evaluation, with two additional measures. A different study by [14] also used the VGG OxfordNet as encoder and RNN-LSTM as decoder. Arabic Flickr8 plus sample of MS COCO data set with Arabized captions have been used. The authors in [2], translated the captions of Arabic Flickr8 data and made it available online. The authors [2] also proposed a model using VGG16 CNN and LSTM as encoder and decoder respectively. They also proposed a base model, which generate English captions, which are then translated to Arabic.

B. *Encoder – Decoder Models with Attention Mechanism*

In [15], the authors proposed three different encoders, utilizing CNNs for feature extraction and single, and/or multiple objects detection. A final hybrid model was proposed with attention mechanism, which is used for detected objects prioritization. They used LSTM with soft attention and beam search were used for the decoder. The data set used was MS COCO, and Flicker30. Unlike other studies, the authors assessed the quality of the generated captions by measuring the semantic similarity between the generated captions and ground truth captions. Another method that used attention is [16], as shown in the next section.

C. *Encoder – Decoder Architectures with Transformers*

Two studies reported their results [16] and [17]. In [16] three models were proposed. The first uses MobileNetV2 network as encoder, LSTM with attention as the decoder. The second was MobileNet V2 (GRU) as encoder, and GRU with attention as decoder. Finally, a transformer-based model was also proposed, where EffeceintNet is used as the encoder, and a transformer based architecture as the decoder. FARASA segmenter has been used for text pre-processing, and BLEU-N scores were reported. Different transformer based models were proposed in [17] which were initialized with AraBERT and GigaBERT pre-trained transformers, then fine-tuned by detecting object tags in images using OSCAR method. Flickr8 and part of MS COCO data sets have been used, and BLEU-N scores are reported.

D. *Analysis of the Related Work*

A comparison of the BLEU-N scores for the studies reviewed in this section can be found in Section IV, Table IV. From that table, it can be seen that the best reported BLEU-N scores are by [12], who used root words to generate captions, followed by [14], then the transformers based models in [16]. It was also noted that transformers achieved minor improvements on the BLEU-N scores. However, the comparison of BLEU-N scores does not provide a concrete conclusion which model is better; as most of the studies did not use a common train/ validation/ test splits. For example, [16] used 90%, 10 % for training, and testing respectively. Also, in [17], the MS COCO images used only for training, Flickr8 test set have been used for testing. Similarly, in [13], 1500, 250, and 250 images have been used for training, validation and testing respectively. Excluding the results by [12], the BLEU-4 results are close, which does not pinpoint the value of a specific architecture over the others. Furthermore, none of the studies reviewed here investigated the effect of pre-processing methods on the BLEU-N scores or the quality of the generated captions.

III. Experimental Methodology

A. *The Dataset*

The data set used in this paper is the Arabic-Flickr8 [2], which is a translated version from the original English Flickr8, using Google Translate. The best 3 translated captions are kept and further edited by native Arabic speakers. For the purpose of training, validation, and testing, Karpathy's data splits [18]; 6000, 1000, 1000 images for training, validation, testing respectively are used. Unlike its English version which has 5 captions for each image, the Arabic Flickr 8 has only 3 captions.

B. *The Model Architecture*

The model used in this paper follows the Encoder- Decoder architecture, with attention mechanism, teacher forcing, and beam search. The selection of the Encoder and Decoder networks is based on their prior excellent performance in computer vision and NLP problems, details as follows:

1) The Encoder: The RESNet-101 [19], is a CNN that has 101 layers, and was chosen -for the first time in Arabic image captioning**-** as the encoder; due to its proven ability to extracts very rich feature set from an image. RESNet stands for Residual Neural Network architecture [19], which is able to overcome the vanishing gradients problem using skip connections. The output of the encoder is passed to the next part of the architecture with its same dimensions.

2) The attention mechanism: The objective of using attention mechanisms [7] in image captioning is to allow focusing on a specific part of the image, while generating the captions. It calculates the weights of different pixels of the encoded images, which are then used by the Decoder. In this paper, soft attention has been used, which is trained in an End-to-End manner using Back-propagation. The soft attention weights are determined by image features and the LSTM previous output.

3) The decoder: An LSTM network is used as the decoder. The LSTM is a variation of the RNNs with additional gates, and is able to overcome the vanishing gradients problem encountered when processing long sequences. Those gates make the RNN decide which tokens should be retained in the memory and which to forget [3]. In the context of image captioning, the decoder looks at different parts of the image; while producing different parts of the output sequence, by weighting different pixels of the output of the encoder. The LSTM cell and hidden states are initialized using the encoded image at the first step; and the encoded image attention weights alongside the decoder weights at each step are computed. The attention weights with the embedding of the token from the previous step, are then concatenated and the LSTM produces the new states.

C. Pre-processing

In this section, the pre-processing steps used are described:

1) Captions' Pre-processing and Tokenization: In this study, Pyarabic [20], which splits image captions into tokens using spaces, and The FARASA segmenter [21] have been used. FARASA [21] is an Arabic word segmenter, which breaks Arabic words into their constituent clitics. For example, the word "wkatabna" (و كتبنا) meaning: "and we wrote" is composed of three clitics "w+katab+na", namely the conjunction article "w" meaning "and" as prefix, the stem "katab" (كتب), possessive pronoun "na" (نا) as suffix. Another example which is very common is the "AL - ال", which corresponds to "the" in English. In the context of image captioning, using FARASA segmenter results in a smaller unique vocabulary size because all different forms of a word are treated as one, but the number of total number of tokens increases, as different suffixes are separated and counted.

2) Image pre-processing: The image pre-processing is kept to the minimum; where all images are resized to 256 (smaller edge) pixels. The center 224x244 pixels was cropped, before transforming them to Pytorch tensors; and normalize using the mean and standard deviation of the IMAGENET dataset.

D. Pre-trained Word Embedding

The Aravec; a pre-trained word embedding have been used to provide richer representations for the image captions. In this study, the skip gram model trained with Wikipedia data have been used [22].

E. The Beam Search

For training and validation, teacher forcing has been used [3], as it makes the model learn the context in more efficient way. At the testing stage, beam search has been used to generate the best captions. Beam search [3] works by finding the top-k words with the highest decoder scores at each step, calculate the additive scores for each of the pairs from current and previous steps and get the best combinations, in each decoding step. In this way, beam search outputs the completed sequences with highest scores. The beam size that resulted in captions with best BLEU-4 score is chosen.

F. Modeling

Two models have been designed to generate image captions; and investigate the effect of the following settings on the quality of the generated captions:

- The use of different text pre-processing in particular, FARASA word segmenter and PyArabic tokenization,
- The use of different Beam Sizes.

1) Model 1: It was decided to start with a base model, as a reference for comparison. So, PyArabic tokenizer, with AraVec pre-trained embedding was used, to compensate for the smaller number of captions for the Arabic Flicr8 data set, compared to its English version.

2) Model 2: In this model, FARASA Segmenter has been used to pre-process the captions, and similar to model 1, AraVec embedding model has been utilized.

For both models, beam search has been used to generate best captions, and they were evaluated on Flickr8 test set, as well as 200 images randomly selected from MS COCO data set, to further test the robustness of our models.

G. Evaluation Methods

1) The BLEU–N Score: The BLEU [23] stands for Bilingual Evaluation Understudy, which is a metric originally proposed to evaluate the quality of machine translation models. As image captioning can be thought of as translation from image features to text describing that image, it has been widely used to evaluate image captioning models. BLEU-1, 2, 3, 4, measures the fraction of n-grams that appear in both the generated and ground truth captions, where n takes the values, 1, 2, 3 and 4.

2) The THUMB 1.0: As human evaluation of the generated captions is still considered the gold standard, the THUMB framework for caption's evaluation is used. THUMB stands for "Transparent Human Benchmark" [24]. THUMB is based on two major scores; namely, precision and recall, which are measured on a scale from 1 to 5. This in addition to penalty scores which are deducted from the average of precision and recall scores, to penalize any problems in the fluency and /or conciseness of the generated captions, as well as any issues concerning the use of inclusive language. The following sections briefly introduce the THUMB 1.0 [24] framework.

a) Precision: As the measure entails, the Precision (P) assess how precise the image is described by the generated caption, which is mainly intended to detect the common failures of the language model part. Precision is measured on a scale from 1 to 5, where 0.1 point is deducted in situations like, minor difference in colors, counts, the caption is not accurate, but do not mainly contradict with image's contents, in addition to other attributes like occasions, locations, etc. [24].

b) Recall: Recall (R) evaluates how good (complete) does the caption describe the image contents; including main objects, their relationships and colors. Therefore, it penalizes the generic, short captions that are usually generated by the majority of image captioning methods. If the image description (caption) is too generic, where different diverse images can be imagined based on that caption, then the recall score tends to be low [24].

c) Penalties: Penalties are given to penalize fluency problems; which assesses the text structure, regardless of the image contents. As most of automatically generated captions do not suffer fluency problems, points are deducted from the average of the precision and recall scores. A penalty of 0.1 is given for grammatical or spellings mistakes, as they are easily corrected. For other more serious problems like duplication, broken sentences, a minimum of 0.5 points are deducted. The Conciseness is also evaluated in this framework, where penalties are given for unnecessarily long, detailed captions, where 0.5 points are deducted. However, as the majority of automatically generated captions tend to be short, this penalty is not very common [24]. The final type of penalties is given on describing humans with terms that deviate from inclusive language**,** which ranges from 0.5 for subjective comments, to 2.0 for more severe problems. A final rule in this framework, is that double penalties should be avoided. If a problem is penalized using precision, it should not be penalized again by recall [24].

IV. Results and Discussion

A. Model 1 Results

Results of model 1 are shown in Table I. From this table, it can be seen that the model achieved BLEU-4 score of 8.29, which is superior to 9 of previously published BLEU-4 scores. This can be attributed to the strong encoder architecture used as well as the use of the attention mechanisms and beam search.

B. Model 2 Results

Results of model 2 are shown in Table II. From this table, it can be seen that the BLEU-N scores have significantly increased for all beam sizes. The best BLEU-4 results were obtained with beam size of 5. The results show that the use of FARASA segmenter significantly improved all the scores. These results are consistent with [12], where root words have been used, which is a similar approach to FARASA pre-processing. This model achieved *the highest BLEU-4 scores* obtained on the Arabic Flickr8 data set so far. Fig. 1 shows the BLEU-4 scores obtained at different beam sizes, for PyArabic, compared to FARASA pre-processing. From this figure, it can be seen that the best results were obtained with beam size 5. This can be attributed to the improvements of the completed sequence of words with beam size of 5, compared to the ground truth captions for the test set.

C. Results on 200 Images from MS COCO

As an additional test of the models quality 200 randomly selected images from MS COCO data set with Arabized captions were used. The results are shown in Fig. 2. From this figure, it can be seen that better results were obtained; again by the model with FARASA segmenter, and beam size of 3, which is the same situation as Flickr8 test set. Fig. 3 compares the performance of the two models on MS COCO data set.

TABLE I. Scores of Model 1 on Arabic Flickr8 with Beam Search

Model	**BLEU-1**	**BLEU-2**	**BLEU-3**	**BLEU- 4**
Model 1 + Beam size 1	39.01	24.45	13.01	7.27
Model 1 + Beam size 3	40.10	25.58	14.28	7.89
Model 1 + Beam size 5	39.10	25.13	13.96	**8.29**

TABLE II. Comparison of BLEU-4 Score for FARASA pre-processing, and Beam Sizes

Model	**BLEU 1**	**BLEU 2**	**BLEU 3**	**BLEU 4**
FARASA + Beam size 1	57.45	43.79	31.86	22.81
FARASA + Beam size 3	59.90	47.40	36.13	26.89
FARASA + Beam size 5	58.71	46.52	35.71	**27.12**

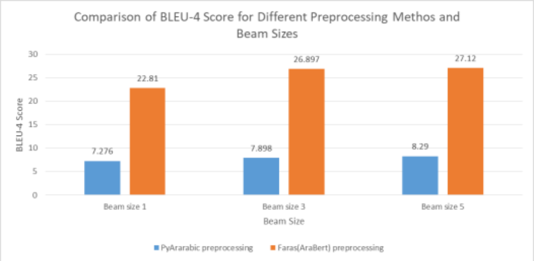


Fig. 1. Comparison of BLEU-4 Score for different Preprocessing Methods and Beam Sizes.

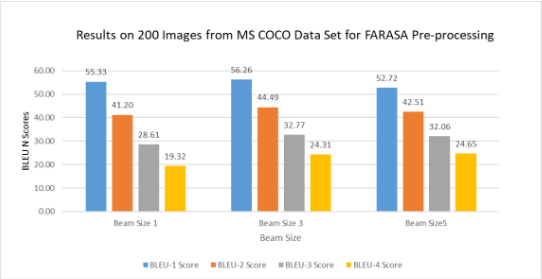


Fig. 2. Comparison of BLEU-4 Score for different Beam Sizes.

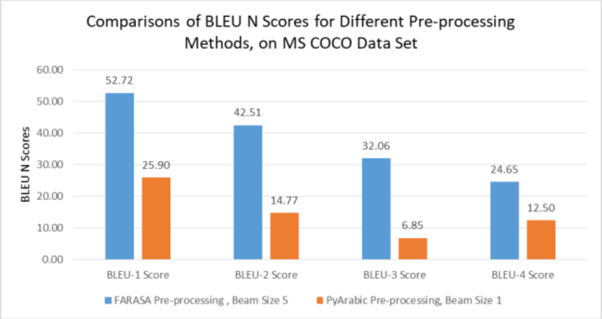


Fig. 3. Comparison of BLEU-N Scores, for Models 1 and 2, on 200 Images from MS COCO.

D. *Results' Comparison with Related Work*

Fig. 4 shows our BLEU-4 results, compared to previously published work, and Table III shows our BLEU-N scores as compared to previously published results on Arabic Flickr8. From this table, it can be seen that the highest BLEU-4 score are achieved by the model trained with FARASA segmenter and beam size of 5. Furthermore the PyArabic model also achieved high BLEU- N scores, which are superior to 9 of related work results. As noted in the related work Section II, most of those studies did not report the data splits they used, others used different splits like [14] and [17], who used parts of both Flickr8 and MS COCO data sets, and [16], who used 90/10 for training and testing, while [17] used combined data set for training, but the testing was only done on part of Flickr8 data.

E. *Qualitative Evaluation of our Results*

As a complementary measure to the obtained BLEU–N scores, it was important to seek qualitative evaluations, to validate the BLEU-N scores results, and show that our models predict accurate and meaningful description for the images. Table IV shows a sample of the captions generated by our models, compared to others previously published captions for same images. As human evaluations are still considered the gold standards to evaluate the quality of machine generated captions, four native Arabic speakers were asked to rank ours, and others captions for each image. Based on the average ranking for the evaluators, they reported that our model's captions have better quality in 16 out the 29 (55%) captions listed in Table IV. In particular, those which are generated with models used PyArabic pre-processing. Other related work methods are better in 8 out of 29 (28%), and both have the same quality in 5 out of the 29 (17%) images.

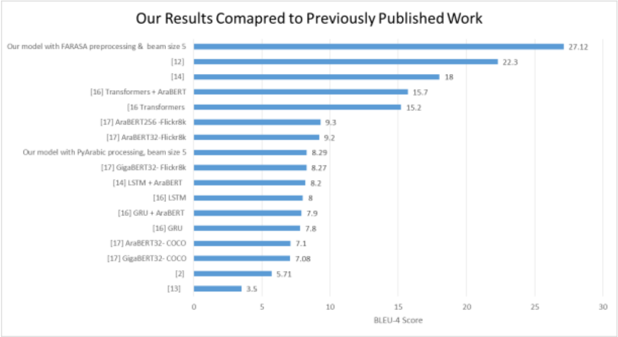


Fig. 4. Our Results Compared to those of Previously Published Methods.

TABLE III. OUR RESULTS AS COMPARED TO PREVIOUSLY PUBLISHED METHODS

Model	**BLEU 1**	**BLEU 2**	**BLEU 3**	**BLEU 4**
[2]	33.18	19.26	10.49	5.71
[16] Transformers + ARABERT	44.30	N/A	N/A	15.70
[16] Transformers	42.70	N/A	N/A	15.20
[16] LSTM + ARABERT	38.30	N/A	N/A	8.20
[16] LSTM	35.10	N/A	N/A	8.0
[16] GRU + ARABERT	37.6	N/A	N/A	7.9
[16] GRU	35.3	N/A	N/A	7.8
[14]	52.00	46.00	34.00	18.0
[12]	65.8	55.9	40.4	22.30
[13]	34.40	15.40	7.60	3.50
[17] AraBERT32-Flickr8k	39.10	24.6	15.0	9.2
[17] AraBERT32- COCO	36.5	22.1	12.9	7.1
[17] AraBERT256 -Flickr8k	38.7	24.4	15.1	9.3
[17] GigaBERT32- Flickr8k	38.6	24.1	14.4	8.27
[17] GigaBERT32- COCO	36.0	21.5	12.4	7.08
Our model with PyArabic & beam size of 3	39.108	25.131	13.962	8.29
Our model with FARASA & beam size 5	58.708	46.523	35.712	**27.12**

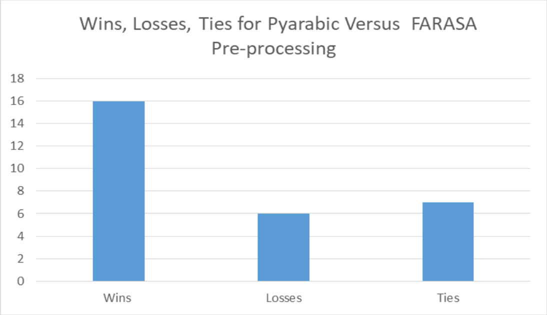


Fig. 5. Summary of THUMB Score Comparison between PyArabic, and FARASA based Models.

To further evaluate and quantify the differences in the captions' quality, THUMB 1.0 [24] was employed, which is utilized for Arabic image captioning for the first time in this paper**.** Therefore, three different native Arabic speakers were asked to use the THUMB framework, and evaluate the captions in Table IV. Based on the average of ratings by the three evaluators; Fig. 5 compares the THUMB scores of our two models. From this figure, it can be seen that PyArabic model has better quality captions, where it obtained 16 wins, 6 losses and 7 ties; compared to FARASA based model. Similarly, Fig. 6 summarizes the comparison results between our models and the related work, showing our models' wins, losses and ties. Again, and similar to the first four evaluators, our models obtained higher THUMB scores, where 15 and 17 wins; 5 and 7 losses; and 9 and 5 ties were obtained by our PyArabic and FARASA based models respectively. Fig. 7 shows detailed comparison between our model's Precision, Recall, and Penalties, compared to those of related work. From this table that our models obtained better scores, but had more penalties, in particular for FARASA based model. Even though

FARASA based model compares favorably to the related work models; however, it has more penalties, due to token repetitions, like example 2, location issues, like example 6, where the caption “little girl in people” should have been “little girl with people” , in Arabic “فتاة صغيرة في الناس” should have been “فتاة صغيرة مع الناس”. Again, in spite of the fact that BLEU-N scores for FARASA based model are much higher than PyArabic based model, it tends to output short, more generic descriptions, compared to PyArabic based models, where the structure and/or the semantics of the captions are better for the latter. The improved results for FARASA based models could be partially attributed to the increase of the number of tokens, like the suffix “AL” "ال", possessive pronoun “na” (نا), “TAA Marboota” “ة”, while number of unique stem words decreases. Another reason is that; due to the tendency of producing short, more generic captions, then the overlap of the N-grams with the ground truth captions in the test set increases, hence the higher BLEU-N scores. The important question here is: should FARASA based models be used as they have higher BLEU-N scores? The answer depends mainly on the nature of the image to be captioned. For example, the captions for images 1, 5, 7, 13 are of same quality for both models. However, if the image scene is busier, then PyArabic based models would be better, like the case of images 2, 3, 6, 18, etc. As Fig. 5 shows, PyArabic based model produces better, stronger sentences semantically; however, the difference of the models scores are minor in most of the cases, which favors the use of FARASA based models. It should be noted here that it is hard to conclude that PyArabic based models are qualitatively better than FARASA based models; as only a sample of 29 images were shown here, which are common between all published work, and obtaining the THUMB scores for more images could reverse the situation, which is likely, as the BLEU-4 score for FARASA based model is much higher. One more thing that could give us a hint, and may help in answering the question; is to be able to understand the internal logic of the model during the caption generation process. This can be achieved by visualizing the attention mechanism weights during caption generation, which is explored in Section V.

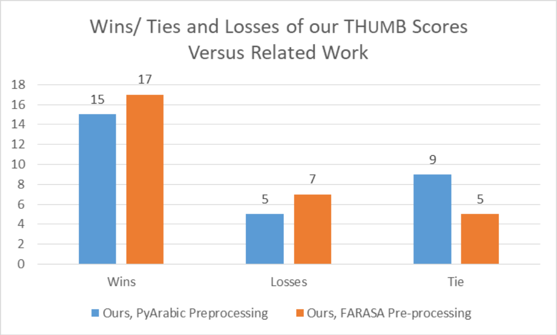


Fig. 6. Summary of THUMB Score for our Two Models, Compared to Related Work Models.

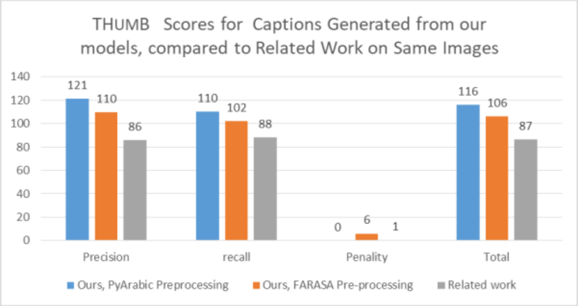
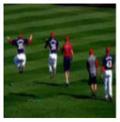
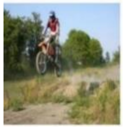
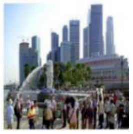
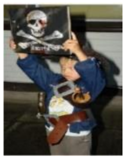
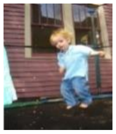
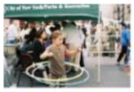
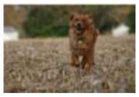
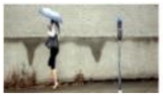
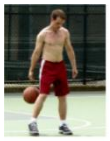
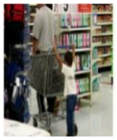
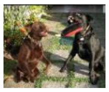
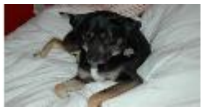
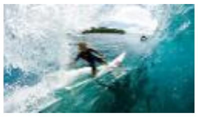
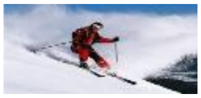
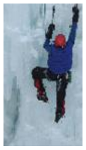
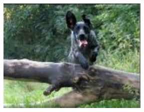
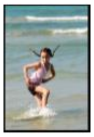
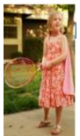
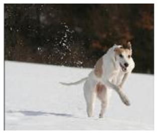
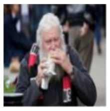
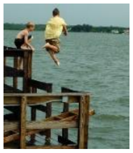
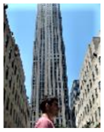
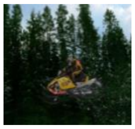
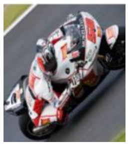
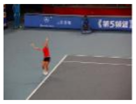
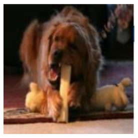
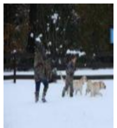
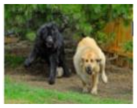
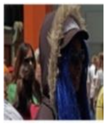


Fig. 7. Detailed THUMB Scores for Our Two Models, Compared to Related Work Models.

TABLE IV. OUR GENERATED CAPTIONS, COMPARED TO PREVIOUS STUDIES FOR THE SAME IMAGES

	Image	*PyArabic based model Captions*	*FARASA based model captions*	*Previous work captions*
1		مجموعة من الرجال يلعبون كرة القدم Group of men playing football	مجموعة من الناس يلعبون كرة القدم Group of people playing football	**[1]** لاعب كرة قدم يرتدى قميص احمر اللون فى الملعب A soccer player wears a red shirt on the field
2		شخص يركب دراجة نارية Person riding a motorcycle	راكب الدراجة الدراجة الدراجة Cyclist bike bike	**[1]** رجل يركب دراجة ترابية Man riding a dirt bike
3		مجموعة من الناس يقفون في الشارع A group of people standing in the street	مجموعة من الناس في الشارع Group of people in the street	**[1]** مجموعة من الناس فى الخارج فى مدينة مزدحمة A group of people outside in a crowded city
4		صبي صغير يرتدي سترة حمراء Little boy wearing a red jacket	رجل يرتدي قبعة صغيرة man wearing beanie	**[16]** صبي صغير يرتدى زى القراصنة يرفع علم القراصنة A little boy in a pirate costume raises a pirate flag

5		صبي صغير يقفز في الهواء Little boy jumping in the air	صبي صغير يقفز في الهواء little boy jumping in the air	**[16]** صبي صغير يقفز في الهواء little boy jumping in the air
6		مجموعة من الناس يلعبون في الهواء A group of people playing in the air	فتاة صغيرة في الناس little girl in people	**[16]** صبي صغير فى قميص ازرق و جينز ازرق Little boy in a blue shirt and blue jeans
7		كلب بني يركض عبر العشب Brown dog running through the grass	كلب بني يركض في العشب Brown dog running in the grass	[16] كلب بني يركض في حقل Brown dog running in a field
8		فتاة صغيرة في قميص أزرق Little girl in blue shirt	امرأة صغيرة في الهواء little woman in the air	**[16]** امرأة فى ثوب السباحة تمشى فى بركة A woman in a bathing suit walking in a pool
9		رجل في قميص أزرق يلعب كرة السلة A man in a blue shirt playing basketball	رجل يرتدي سترة السلة السلة A man wearing a basketball jacket the basket the basket	**[16]** صيي صغير يلعب كرة السلة فى ملعب رياضى Little boy playing basketball in the sports field
10		صبي صغير يقفز في الشارع Little boy jumping the street	امرأة صغيرة في الشارع little woman in the street	**[16]** رجل يعزف على الجيتار فى الشارع A man playing guitar in the street
11		كلب أسود يركض على العشب Black dog running on the grass	اثنين من الكلاب يلعبون في العشب Two dogs playing in the grass	**[2]** كلب بنى يقف فى الماء Brown dog standing in the water
12		كلب أسود يركض في الثلج Black dog running in the snow	اثنين من الكلاب يلعبون في الثلج Two dogs playing in the snow	**[14]** A black and white كلب أسود وأبيض يقفز على سجادة dog jumps on a carpet
13		رجل يركب الأمواج Man surfing	راكب الأمواج في الماء Surfer in the water	**[14].** رجل يمارس رياضة ركوب الأمواج Man surfing
14		رجل يرتدي سترة حمراء على لوح التزلج على الجليد A man wearing a red jacket on a snowboard	المتزلجان في الثلج Snow skaters	**[14]** رجل يرتدي خوذة حمراء قيف على تلة ثلجية A man wearing a red helmet stood on a snow hill
15		صبي صغير يقفز في الثلج Little boy jumping in the snow	المتزلج في الثلج snow skater	**[2]** صبى فى سترة حمراء يلعب فى الماء A boy in a red jacket playing in the water
16		كلب أسود يقفز في الهواء Black dog jumping in the air	كلب أسود يقفز في الهواء Black dog jumping in the air	**[2]** كلب اسود يقفز فى الهواء Black dog jumping in the air
17		فتاة صغيرة في الماء Little girl in the water	فتاة صغيرة في الماء على الشاطئ little girl in the water on the beach	**[2]** صبي فى ثوب سباحة يلعب فى الماء Boy in a bathing suit playing in the water

18		فتاة صغيرة في قميص أحمر في حقل Little girl in a red shirt in a field	فتاة صغيرة في العشب little girl in the grass	**[13]** فتاة صغيرة ترتدي فستان ملون تحمل كوب bear cup little girl wearing a colorful dress
19		كلب أبيض يركض في الثلج A white dog running on the snow	كلب أبيض يركض في الثلج A white dog running on the snow	**[2]** كلب أبيض يركض في الثلج White dog running in the snow
20		رجل في قميص أزرق Man in blue shirt	امرأة ترتدي سترة وامرأة Woman wearing a jacket and woman	**[1]** تحمل الزهور من باقة ا مرأة فى سترة حمراء Woman in red jacket carrying flowers from a bouquet
21		رجل يجلس على لوح التزلج على الشاطئ A man sitting on a surfboard on the beach	رجل يقفز في الماء Man jumping in water	**[16]** صبيان يستعدان للقفز من رصيف يقع على جسم كبير فى الماء Boys preparing to jump from a pier that falls on a large body in the water
22		رجل يقف على مقعد في الشارع Man standing on bench in the street	امرأ ة من الناس في الشارع Woman of people in the street	**[2]** مجموعة من الناس يحملون المشروبات و يشيرون الى الكاميرا A group of people carrying drinks and pointing at the camera
23		شخص يركب دراجة على الجليد person riding a bicycle on ice	شخصان في الثلج Two people in the snow	**[13]** رجل يرتدي خوذة زرقاء اللون يقفز فوق لافتة تقول Man wearing blue helmet color jumps over a sign that says
24		رجل يركب دراجة نارية Man riding a motorcycle	مجموعة من الدراجة النارية set of motorcycle	**[13]** رجل يرتدي خوذة حمراء يقود دراجة نارية في الهواء Man wearing red helmet driving a motorcycle in the air
25		رجل في قميص أحمر يقفز في الهواء A man in a red shirt jumps in the air	رجل يقفز على الهواء man jumping on air	**[15]** امرأة تحمل مضربا فوق ملعب تنس Woman holding a racket on a tennis court
26		كلب بني يركض على العشب Brown dog running on the grass	الكلب البني والبني Brown dog brown dog	**[13]** كلب بني و اسود اللون يقفز فوق سياج ابيض و ابيض Brown and black dog jumping Over a white and white fence
27		اثنين من الكلاب يلعبون في الثلج Two dogs playing in the snow	ثلاثة من الكلاب في الثلج Three dogs in the snow	**[13]** كلب اسود و كلب بني اللون في حقل عشبى Black dog and brown dog in a grassy field
28		كلب أسود يركض عبر العشب Black dog running across the grass	اثنين من الكلاب يلعبون في العشب Two dogs playing in the grass	[13] كلب اسود و كلب بني اللون في حقل عشبي. Black dog and brown dog in grass field
29		اثنين من الناس في الهواء Two people in the air	مجموعة من الناس في الشارع group of people on the street	[16] امرأة فى سترة برتقالية تتحدث على هاتفها المحمول A woman in an orange jacket talking on her mobile phone

V. Attention Mechanism and Pre-processing Methods

Motivated by the results obtained in Section IV, where the use of PyArabic and FARASA segmenter lead to different BLEU-N; as well as THUMB scores, we thought it would be interesting to visualize the sequence of the tokens generated with images' attended parts by each model. Therefore, attention scores visualization is used to investigating the effect of different text pre-processing methods on the generated captions. Put it another way; it is required to investigate whether the attention mechanism attends the image salient features the same way during word sequence generation; for different pre-processing methods.

Fig. 8 to 15 show the attention visualization for four different images and their captions, while the alignment of tokens is shown with the images' attended parts during sequence generation. Interestingly, it is noticed from those figures, that the attention mechanism behavior is somehow different in FARAS based model compared to PyArabic based model. Take for example Fig. 8 and 9, which visualize the attention by FARASA and PyArabic models respectively. Even though the length of the sequence is shorter in Fig. 8 (FARASA based model), it can be noticed that the model perfectly attends the token "فتاة"," Girl" , which spans images 2,3,4, then "in" "فى"in 5 , then " the water" "ماء+ ال"6,7,8, which made the attended image features very prominent. This is not the case in Fig. 9, where an additional word "water""ماء" is not clearly visualized like Fig. 8. Looking at Fig. 10 and 11, it can be seen that the attention mechanism better attends the image's salient features in case of FARASA based model, where the attention shows the dogs and the snow clearly, which is not the case in Fig. 11 for PyArabic. One explanation is that FARASA pre-processing outputs larger number of tokens for the same image region, therefore, higher attention scores.

In addition to the above, attention visualization is also useful in detecting a model's failure, even if it is not very clear in the caption. Consider for example Fig. 12 and 13, where the generated caption is " child in the air", while the attention is attending the cat face, which is the same for PyArabic based model, "Man is sitting on a chair", still the attention is on the cat's face. This is a clear failure of both models, which may not be noticed in case of PyArabic model, as there is a man sitting there in the image. From those examples, it can be concluded that FARASA based model better attends the image's salient features during caption generation, compared to PyArabic model. A third, and more interesting example is shown in Fig. 14 and 15, where an images composed of two individual images, for two different scenes are stacked together. For Fig. 14, the generated caption is "three people in the water", and the attention shows that the dog is counted as one of the three persons in the image, and the water is clearly attended as well. This is not the case for Fig. 15, with the caption "Man Surfing", where one person is attended, while the surfing token is aligned with the dog, which is a clear failure of the model.

From those examples, it can be concluded that FARASA segmenter improves the attention mechanism behavior, which explains the high BLEU-N scores of its model. However, more examples needs to be looked at, to confirm such conclusion.

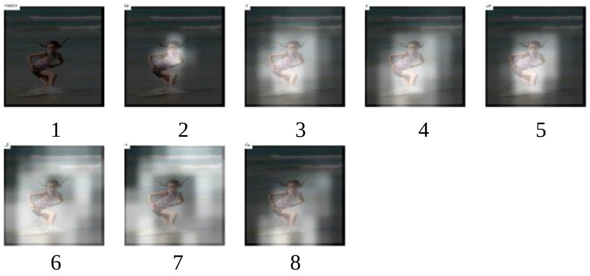


Fig. 8. Attention Weights Visualization for FARASA based Model: The Caption is "Girl in the Water".

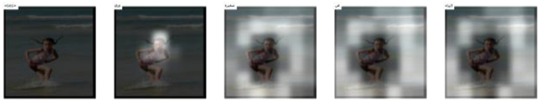


Fig. 9. Attention Weights Visualization for PyArabic based Model. The Caption is "Little Girl in the Water."

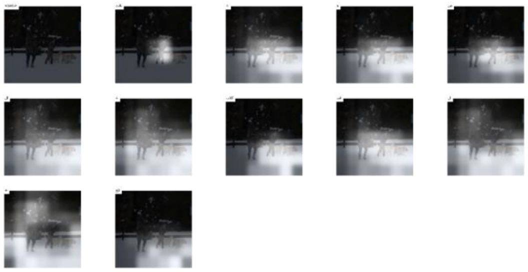


Fig. 10. Attention Visualization for FARASA based Model. The Caption is: "Three Dogs in the Snow".

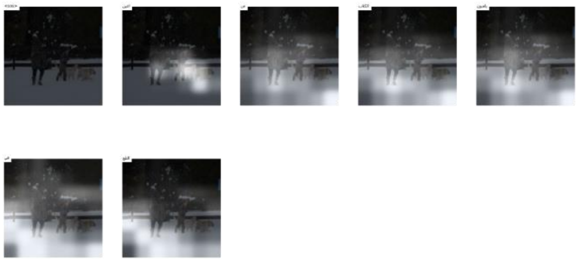


Fig. 11. Attention Visualization for PyArabic based Model. The Caption is: "Two Dogs Playing in the Snow".

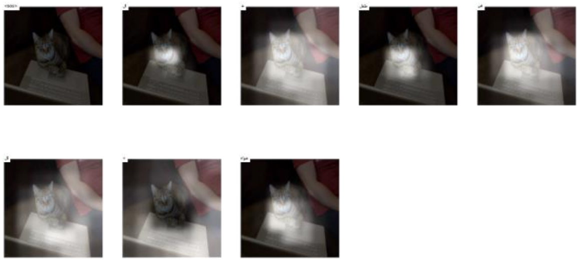


Fig. 12. Attention Visualization for FARASA based Model. The Caption is: "Child in the Air".

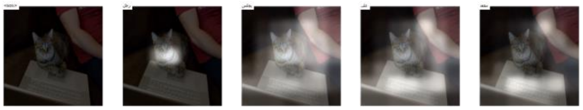


Fig. 13. Attention Visualization for PyArabic based Model. The Caption is: "Man Sitting on Chair".

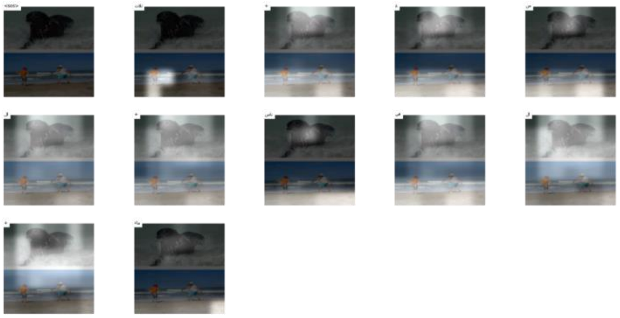


Fig. 14. Attention Visualization with FARASA based Model. The Caption is: "Three People in Water".

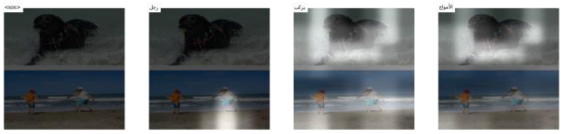


Fig. 15. Attention Visualization for PyArabic based Model. The Caption is: "Man Surfing".

VI. Conclusion and Future Work

In this paper, an efficient model for Arabic image captioning is proposed, and the effect using beam search, as well as the pre-processing on the testing BLEU-4 score is investigated. The model with FARASA segmenter achieved the state of the art BLEU-4 score. The results are also consistent with the results of the model which used root words to generate Arabic captions. In addition to the BLEU-N scores, the generated captions were also qualitatively evaluated by two teams of Arabic native speakers, the first team the captions, while the other used the "THUMB" framework for evaluation. The generated captions using two different text pre-processing models achieved the best THUMB scores, where the model with PyArabic pre-processing showed better results on the sample used. Another interesting finding in this paper is that the different text pre-processing methods influence the attention mechanism, where FARASA based model showed better attention visualization for the used samples. The generated captions also compares favorably to all previous related work, quantitatively, and qualitatively. The paper also show that the choice of the right architecture, with the right pre-processing of Arabic text and the use of beam search can significantly improve the quality of the generated captions.

From the work done in the area of Arabic image captioning including this study, it can be concluded that the use of transformers did neither significantly improve the BLEU-N results, nor the use of larger data sets in training.

As a direction of future research, more efficient models can be investigated to improve the obtained results in this area, utilizing Generative Adversarial Networks. Another direction is to propose new text pre-processing methods and additional evaluation methods to cope with morphologically rich languages like the Arabic language.

References

[1] M. Cheikh and M. Zrigui, "Active Learning Based Framework for Image Captioning Corpus Creation," in International Conference on Learning and Intelligent Optimization, Cham, 2020: Springer International Publishing, in Learning and Intelligent Optimization, pp. 128-142.

[2] O. ElJundi, M. Dhaybi, K. Mokadam, H. Hajj, and D. Asmar, "Resources and End-to-End Neural Network Models for Arabic Image Captioning," in 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) 2020, vol. 5, pp. 233-241.

[3] M. Stefanini, M. Cornia, L. Baraldi, Silvia Cascianelli, G. Fiameni, and R. Cucchiara, "From Show to Tell: A Survey on Deep Learning-based Image Captioning," arXiv:2107.06912v3, 2021.

[4] H. Wang, Yue Zhang, and X. Yu, "An Overview of Image Caption Generation Methods," Computational Intelligence and Neuroscience, vol. 2020, p. 13, 2020.

[5] A. Pal, S. Kar, A. Taneja, and V. K. Jadoun, "Image Captioning and Comparison of Different Encoders," Journal of Physics: Conference Series, vol. 1478, no. 012004, 2020.

[6] R. Staniute and D. Šešok, "A Systematic Literature Review on Image Captioning," Applied Sciences, vol. 9, no. 2024, p. 20, 2019.

[7] Z. Zohourianshahzadi and J. K. Kalita, "Neural Attention for Image Captioning: Review of Outstanding Methods," Artificial Intelligence Review, 2021.

[8] A. Elhagry and K. Kadaoui, "A Thorough Review on Recent Deep Learning Methodologies for Image Captioning," arxiv.2107.13114, 2021.

[9] S. Li, Z. Tao, K. Li, and Y. Fu, "Visual to Text: Survey of Image and Video Captioning," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 3, no. 4, pp. 297-312, 2019.

[10] T. Ghandi, H. Pourreza, and H. Mahyar, "DEEP LEARNING APPROACHES ON IMAGE CAPTIONING: A REVIEW " arXiv:2201.12944, 2022.

[11] S. AMIRIAN, K. RASHEED, T. R. TAHA, and H. R. ARABNIA, "Automatic Image and Video Caption Generation With Deep Learning: A Concise Review and Algorithmic Overlap," IEEE Access, vol. 8, pp. 218386- 218400, 2020.

[12] V. Jindal, "Generating Image Captions in Arabic Using Root-Word Based Recurrent Neural Networks and Deep Neural Networks," in The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, Louisiana, USA., 2018: Association for the Advancement of Artificial Intelligence, pp. 8093-8094.

[13] R. Mualla and J. Alkheir, "Development of an Arabic Image Description System," International Journal of Computer Science Trends and Technology (IJCST) vol. 6, no. 3, pp. 205-213, 2018.

[14] H. A. Al-Muzaini, T. N. Al-yahya, and H. Benhidour, "Automatic Arabic image captioning using RNN-LSTM-based language model and CNN," International Journal of Advanced Computer Science and Applications, vol. 9, no. 6, pp. 67-73, 2018.

[15] I. Afyouni, I. Azhar, and A. Elnagara, "AraCap: A hybrid deep learning architecture for Arabic Image Captioning," Procedia Computer Science, vol. 189, pp. 382-389, 2021.

[16] S. M. Sabri, "ARABIC IMAGE CAPTIONING USING DEEP LEARNING WITH ATTENTION," Masters, University of GeorgiaProQuest Dissertations, ATHENS, GEORGIA, 2021.

[17] J. Emami, "Arabic Image Captioning using Pre-training of Deep Bidirectional Transformers," Masters, Computer Science, LUND UNIVERSITY, 2022.

[18] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3128-3137.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in IEEE conference on computer vision and pattern recognition, 2016: IEEE, pp. 770-778.

[20] Pyarabic, An Arabic language library for Python. (2010). [Online]. Available: https://pypi.python.org/pypi/pyarabic/.

[21] K. Darwish and H. Mubarak, "Farasa: A New Fast and Accurate ArabicWord Segmenter," in LREC 2016, Tenth International Conference on Language Resources and Evaluation, Slovenia, 2016, pp. 1070-1074.

[22] A. B. Soliman, K. Eissa, and S. R.El-Beltagy, "AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP," Procedia Computer Science, vol. 117, pp. 256-265, 2017.

[23] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation," in 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 311-318.

[24] J. Kasai et al., "Transparent Human Evaluation for Image Captioning," arXiv:2111.08940v2, vol. cs.CL, May 2022.

Identification of Human Sperm based on Morphology Using the You Only Look Once Version 4 Algorithm

Aristoteles[1], Admi Syarif[2]*, Sutyarso[3], Favorisen R. Lumbanraja[4], Arbi Hidayatullah[5]
Doctoral Program of Mathematics and Natural Sciences, Lampung University[1]
Department of Computer Science, Faculty of Mathematics and Natural Sciences[1, 2, 4, 5]
Lampung University, Bandar Lampung, Indonesia[1, 2, 4, 5]
Department of Biology, Faculty of Mathematics and Natural Sciences[3]
Lampung University, Bandar Lampung, Indonesia[3]

***Abstract*—Infertility is a crucial reproductive problem experienced by both men and women. Infertility is the inability to get pregnant within one year of sexual intercourse. This study focuses on infertility in men. Many causes that can cause infertility in men including sperm quality. Currently, identification of human sperm is still done manually by observing the sperm with the help of humans through a microscope, so it requires time and high costs. Therefore, high technology is needed to determine sperm quality in the form of deep learning technology based on video. Deep learning algorithms support this research in identifying human sperm cells. So deep learning can help detect sperm video automatically in the process of evaluating sperm cells to determine infertility. We use deep learning technology to identify sperm using the You Only Look Once version 4 (YOLOv4) algorithm. Purpose of this study was to analyze the level of accuracy of the YOLOv4 algorithm. The dataset used is sourced from a VISEM dataset of 85 videos. The results obtained are 90.31% AP (Average Precision) for sperm objects and 68.19% AP (Average Precision) for non-sperm objects, then for the results of the training obtained by the model 79.58% mAP (Mean Average Precision). Our research show result about identification of human sperm using YOLOv4. The results obtained by the YOLOv4 model can identify sperm and non-sperm objects. The output on the YOLOv4 model is able to identify objects in the test data in the form of video and image.**

Keywords—Classification; deep learning; identification; sperm; sperm head; you only look once version 4

I. Introduction

In the last two decades, the reproductive problem in men that has received much attention is infertility. Infertility is caused by many things, one of which is abnormalities in sperm morphology. Morphological abnormalities experienced by sperm include thin heads, amorphous heads, or bent or asymmetrical neck is of little clinical use [1]. Many studies have reported some analytical disturbances of sperm morphology tests in the details of the sperm sections that were carried out manually. Several studies related to technology to support the diagnosis of infertility in sperm include Computer Assisted Sperm Analysis (CASA), Automatic Assessment of Biochemical Markers of Seminal Plasma, Histopathology Assessment [2]. The technology development that has been carried out still requires further development to get better results, in order to be able to get accurately analyze and identify infertility problems. Currently, there are many studies that predict the cases of infertility. The method that is commonly implemented is the observation method from patient medical record data at a hospital [3]. A number of studies have shown that the factors that cause infertility include age, smoking habits, marijuana use, heroin, hormone disorders, and immunological disorders [4]. These factors can increase the risk of having abnormal sperm so that it becomes infertility.

Identification of sperm is still mostly done manually by humans by observing directly with the aid of a microscope. This requires a high time and cost. To overcome these problems needed a high technology, it is using deep learning. Research on deep learning has been widely carried out [5, 6, 7, 8, 9].

Convolutional Neural Network (CNN) is a development of deep learning that has been developing since 2012. This method can identify sperm morphology accurately based on images [5]. Accurate identification results are influenced by the quality of full image data or with large pixel sizes, so that accuracy and detection performance can be more optimal [6]. In addition to analyzing the morphology of the sperm, CNN can identify sperm based on motility. The dataset used is video [7]. The best results were obtained at the Mean Average Error (MAE) which was 8.786. This shows that the prediction of sperm motility is a fast and consistent process [8]. In the study [9] used Region Based Convolutional Neural Network (R-CNN) to evaluate sperm head motility in video data [7]. The results obtained in this study were 91% and MAE was 2.92.

Research [8, 9] has not been able to identify sperm and non-sperm objects through video. Therefore, our research carried out the latest and most updated breakthrough in the form of YOLO (You Only Look Once). The YOLO algorithm is a floating of CNN which functions as object detection in multiple images [10]. The YOLO algorithm is a more efficient method than object detection algorithms in other machine learning, because it makes everyone can use a 1080 Ti or 2080 Ti GPU to train a super-fast and accurate object detector [11]. Therefore, it is necessary to build a model using the YOLO method to identify sperm and non-sperm based on video.

*Corresponding Author.

II. Materials and Methods

The process conducted in this study is illustrated in Fig. 1.

Fig. 1 shows an illustration of the workflow this research with many steps. The first step is to collect images for the dataset, then resize the image and annotate it based on the object class. The second step is split data to generate train, validation, and test data. The next step is training data based on predetermined hyperparameters. The last step is to evaluate the model from the results of training data and data testing.

A. Dataset

The data used is from Simula Open Dataset with the address https://datasets.simula.no/visem/. This VISEM dataset is a multi-modal dataset that contains data sources such as videos, biological analysis data, and participant data, however in its use for this study only data in the form of videos are used as research datasets. The VISEM dataset contains 85 video recordings of anonymous data from 85 different donor participants with the AVI (Audio Video Interleave) extension and the resolution of each video is 640 x 480 pixels with 50 fps frame rate. Based on the video, 1330 images are produced which will be processed. Fig. 2 shows a piece frame of microscopic video the VISEM dataset.

B. Annotation Data

In this step, two object annotation classes are given for the image to be trained. The classes created in this annotation are sperm and non-sperm. Each frame is annotated in the form of bounding boxes on the morphology of the sperm head and an object that is not sperm. The results of annotations have been made for sperm class is 105.465 bounding boxes. While for the non-sperm class, 22.425 bounding boxes have been annotated. For this annotation use Yolo_mark which is sourced from GitHub https://github.com/AlexeyAB/Yolo_mark.git. The following display of the data annotation can be seen in Fig. 3.

C. Deep Learning

Deep Learning is a branch of machine learning that is inspired by the human cortex by applying an artificial neural network with many hidden layers [12]. There are many types of Deep Learning, such as Deep Auto Encoder, Deep Belief Nets, Convolutional Neural Network, and others. Deep Learning can solve computer difficulties in understanding the meaning of raw input data that is by breaking the desired complex mapping into a series of nested simple mappings.

D. Convolutional Neural Network (CNN)

Convolutional Neural Network is a subdivision of a Deep Learning algorithm used in computer vision to solve certain cases or problems, such as classifying and detecting objects in images, photos, or videos [13]. The characteristics of CNN have a 3D arrangement of neurons (height, width, and depth). The illustration of CNN architecture can be seen in Fig. 4.

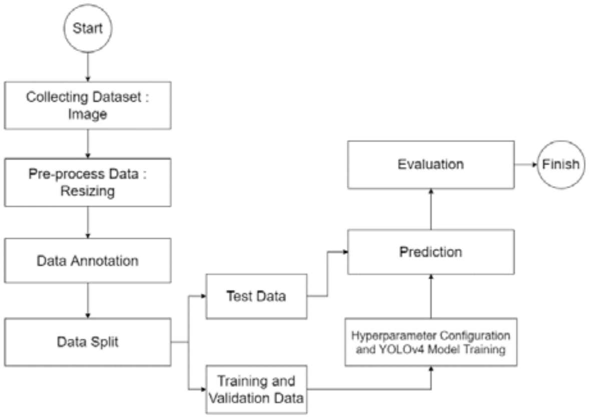

Start
Collecting Dataset : Image
Pre-process Data : Resizing
Data Annotation
Data Split
Test Data
Training and Validation Data
Hyperparameter Configuration and YOLOv4 Model Training
Prediction
Evaluation
Finish

Fig. 1. Research Workflow View.

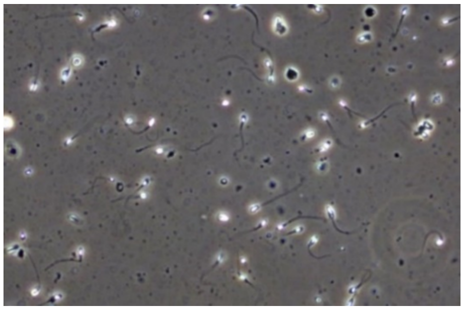


Fig. 2. Frame from Microscopic Video of VISEM Dataset.

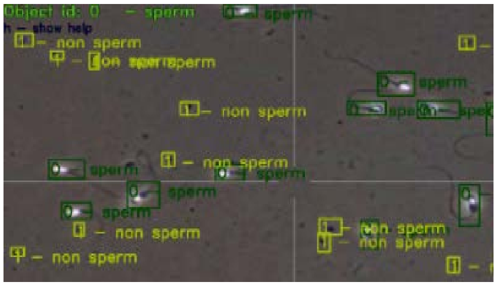


Fig. 3. Display Data Annotation of Image Extraction VISEM Dataset.

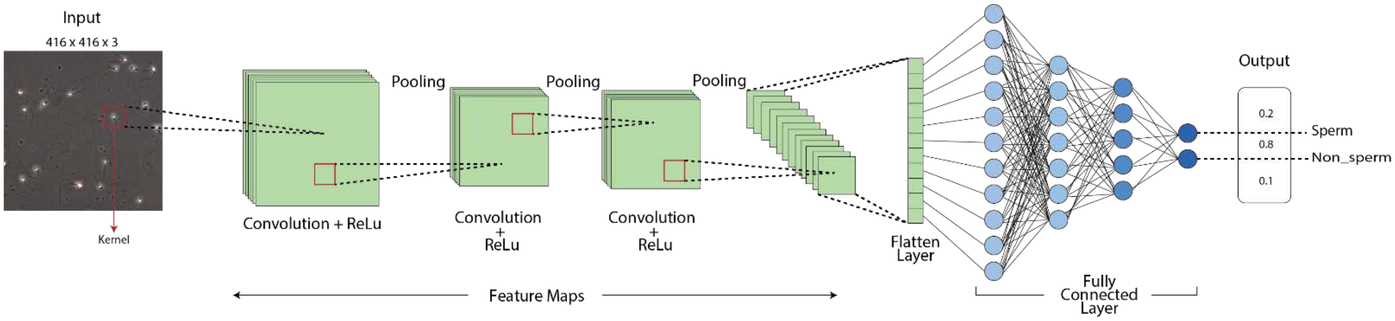


Fig. 4. The Architecture of Convolutional Neural network (CNN) [13].

Fig. 4 shows processes to receive image prediction from CNN. Here we can see the CNN processes are input image, convolution, ReLu, pooling, and fully connected layer.

1) Convolution: The mathematical definition of convolution is the total number of multiplications between the corresponding elements (having the same coordinates) in two matrices or two vectors [14]. Convolution can also be defined as the process of multiplying an image using an external mask or sub-windows to create a new image.

2) Pooling: The pooling layer is a layer that has the function to reduce the spatial size of the convolution process. So as to reduce the computational resources required to process data by reducing the dimensions of the feature map. This pooling layer can make the model training process more effective because the pooling layer is the dominant feature extraction [15].

3) Rectified Linear Units (ReLU): ReLU is the part of the linear function code that removes the negative part to zero and keeps the positive part of the convolution result. Many studies have shown that ReLU outperforms the sigmoid activation function and empirical ground [16]. ReLu activation function is defined as:

$$a_{i,j,k} = max(z_{i,j,k}, 0) \tag{1}$$

Input of the activation function is $z_{i,j,k}$ at location (i,j) on the k-th channel. Simply put, ReLU outputs 0 when $a_{i,j,k} < 0$, and otherwise, it outputs a linier functions when $a_{i,j,k} \geq 0$. Fig. 5 is visual representation of ReLU activation function.

4) Fully connected layer: Fully Connected Layer is a layer that is fully connected; this layer of neurons is connected directly to other neurons by two adjacent layers without being connected to any layer [17]. The Fully Connected Layer processes the output of the final pooling or convolutional layer, which has been flattened. The results of this process will then be continued using the softmax function to get the probability of the input being in a certain class [18].

E. *You Only Look Once (YOLO)*

You Only Look Once or YOLO is a new approach to object detection and development of CNN. YOLO differs from previous research in that detecting an object reuses its classifier; instead YOLO frames object detection as a regression problem with spatially separated bounding boxes and associated class probabilities [10].

Fig. 6 illustrates the YOLO process in making the input image into an image that has been given a bounding box. The first step is the input image will be resized, and then the second step runs a convolutional neural network, after that it does non-max suppression and produces an image that has been identified with a bounding box.

YOLO is implemented as a convolutional neural network. This architecture is inspired by the GoogleNet model for image classification. The YOLO network has 24 convolution layers followed by 2 fully connected layers. Simply uses 1×1 reduction layer followed by 3×3 convolutional layers. The prediction of the final output of this YOLO network is a tensor of 7×7×30. Fig. 7 shows the YOLO architecture in processing image predictions.

There are several versions YOLO of the development of research that has been carried out, in this study using YOLOv4. YOLOv4 overcomes this problem by creating a CNN that operates in real-time on a conventional GPU, and requires only one conventional GPU for training. The purpose of YOLOv4 is to design the speed of operation of the object detector in producing systems and optimization for parallel computing. YOLOv4 hopes that the designed object can be easily trained and used.

Modern detectors usually consist of two parts consisting of a backbone and a head, for a backbone that has been trained previously with ImageNet. Head used to predict the class and bounding boxes of the object. For detectors running on the GPU platform, the backbone used can be VGG, ResNet, ResNetXt, or DenseNet. For detectors running on the CPU platform, the backbone used is SqueezeNet, MobileNet, or ShuffleNet. The development of object detectors in recent years often inserts several layers between the backbone and the head, these layers are usually used to collect feature maps from different stages. We can call this layer the neck object detector. In general, the neck consists of several bottom-up paths and topdown paths. Networks equipped with this mechanism include the Feature Pyramid Network (FPN), Path Aggregation Network (PAN), BiFPN, and NAS-FPN [11].

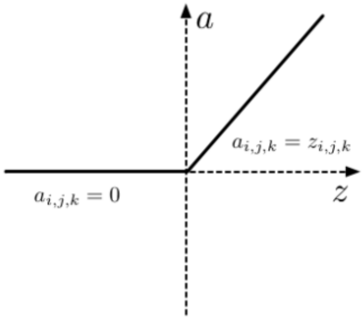


Fig. 5. Display Representation of ReLU Activation Function [16].

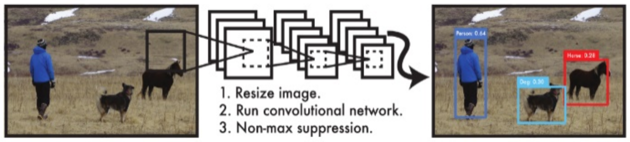


Fig. 6. Processing Images with YOLO Detection System [10].

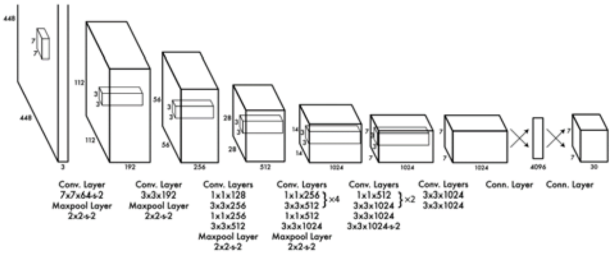


Fig. 7. Display YOLO Architecture [9].

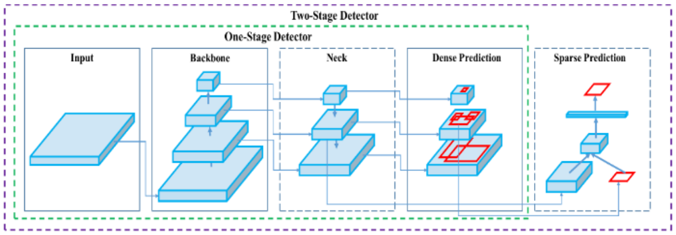
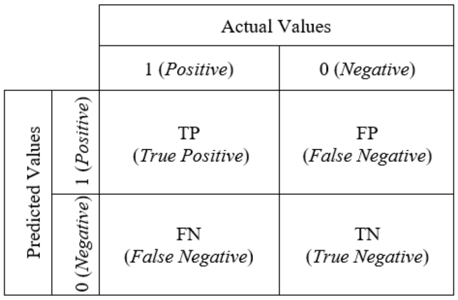


Fig. 8. Structure Modern Detector of YOLOv4 [10].

Based on Fig. 8, YOLO research uses modern detectors, so the researchers make several terms; there are Bag of Freebies (BoF) and Bag of Specials (BoS). The definition of the Bag of Freebies (BoF) is that researchers can perform an optimization to produce better accuracy and not increase inference costs by using training methods. The BoF used for the backbone are DropBlockRegularization, Class Label Smoothing, and CutMix and Mosaic Data Augmentation. BoF for this backbone is useful for increasing the variability in the input image so that the model built has a higher quality for images obtained from different environments. The definition for Bag of Special (BoS) is a set of plugin modules and post-processing methods that only increase inference costs by a small amount, however can significantly improve accuracy in object detection. The BaS used in the backbone include Mish Activation, Cross-Stage Partial Connections (CSP), Multi-Input Weight Residual Connection (MiWRC) [12].

F. *Confusion Matrix*

Confusion Matrix is a performance measurement or performance in solving machine learning classification problems, where the output results can be in the form of two or more classes. Confusion matrix is a predictive analysis tool that displays and compares the actual value with the predicted model value. Prediction models that can be used to get the results of the evaluation matrix are Accuracy, Precision, Recall and F1 Score [19]. This confusion matrix is shown in Fig. 9.



		Actual Values	
		1 (*Positive*)	0 (*Negative*)
Predicted Values	1 (*Positive*)	TP (*True Positive*)	FP (*False Negative*)
	0 (*Negative*)	FN (*False Negative*)	TN (*True Negative*)

Fig. 9. Display Component of Confusion Matrix.

TP (True Positive): The amount of data that is positive and is predicted to be true as positive.

FP (False Positive): The amount of data that is negative however is predicted to be positive.

FN (False Negative): The amount of data that is positive however is predicted to be negative.

TN (True Negative): The amount of data that has a negative value with a correct prediction as negative.

1) Accuracy: Accuracy is a ratio of selected relevant objects to all selected objects. Accuracy can also be defined as a comparison of an object that is correctly identified with the total number of existing objects and the error rate is an object that is identified incorrectly with the total number of existing objects [20].

$$Accuracy = \frac{TP+TN}{TP+TN+FF+FN} \times 100\% \qquad (2)$$

2) Precision: Precision is a level of accuracy of information desired by the user with the prediction results given by the model or system [21].

$$Precision = \frac{TP}{TP+FP} \times 100\% \qquad (3)$$

3) Recall: Recall is the ratio of the number of objects that are detected correctly or True Positive compared to all positive data, recall that has a high value means that the system or model created can classify object classes correctly [22].

$$Recall = \frac{TP}{TP+FN} \times 100\% \qquad (4)$$

4) F1 Score: F1 Score or called the harmonic mean, is a picture of the relative influence between precision and recall [23].

$$F1\ Score = 2 \times \frac{Recall \times Presisi}{Recall + Presisi} \qquad (5)$$

G. *mAP (Mean Average Precision)*

Mean Average Precision or mAP is used as the work evaluation value of the object detection model. Mean Average Precision measures the performance level of the file weights resulting from training mode [24]. The solid mAP equation can be seen in the following equation.

$$mAP = \frac{1}{c}\sum_{t=1}^{c} AP_i \qquad (6)$$

III. Result and Discussion

A. *Train Model Results*

This study uses 3 types of dataset distribution in terms of finding the best level of accuracy. The first division structure is 80% train data, 10% validation data, and 10% test data. The next data division is 70% train data, 25% validation data, and 5% test data. The final data distribution is 60% train data, 20% validation data, and 20% test data. Based on the 3 types of split data, it produces different accuracy values, however the difference in the accuracy values obtained is not too significant. The number of iterations carried out during the training of this model is 6000 iterations. The learning rate was chosen for hyperparameter in model YOLOv4 sperm detection: 0.002, 0.0002, and 0.00002 [9]. The following is a table of accuracy results obtained from training data.

In Table I, testing of all scenarios uses a hyperparameter learning rate of 0.002. We can see that the AP results of each object being trained have the greatest results in two different scenarios. The biggest AP result for sperm objects is in the second scenario with a value of 90.31%. As for the non-sperm object, the biggest result is in the first scenario with a value of

68.35%. The biggest mAP result is in the first scenario of 78.81%.

Table II is a test using a learning rate of 0.0002 from three scenarios. The results obtained from this test are for the biggest AP of the two objects in the second scenario. The AP obtained for sperm objects is 90.37%, while the AP for non-sperm objects is 68.78%. The biggest mAP result is in the second scenario of 79.58%.

Table III uses a learning rate of 0.00002 to get AP results in each scenario. In this training, the biggest AP value for sperm objects is found in the third scenario, with an AP value of 88.42%. As for the AP value of non-sperm objects, the biggest AP result is in the second scenario of 64.40%. The biggest mAP result is in the second scenario of 76.11%.

TABLE I. ACCURACY RESULTS OBTAINED FROM DATA TRAINING WITH A LEARNING RATE OF 0.002

No	**Data Composition**	**AP**		**mAP**
		Sperm	***Non Sperm***	
1	Train 80%, Validation 10%, Test 10%	89.51%	68.13%	78.81%
2	Train 70%, Validation 25%, Test 5%	90.31%	65.35%	77.83%
3	Train 60%, Validation 20%, Test 20%	88.52%	65.24%	76.88%

TABLE II. ACCURACY RESULTS OBTAINED FROM DATA TRAINING WITH A LEARNING RATE OF 0.0002

No	**Data Composition**	**AP**		**mAP**
		Sperm	***Non Sperm***	
1	Train 80%, Validation 10%, Test 10%	89.66%	67.99%	78.83%
2	Train 70%, Validation 25%, Test 5%	90.37%	68.78%	79.58%
3	Train 60%, Validation 20%, Test 20%	89.86%	67.42%	78.64%

TABLE III. ACCURACY RESULTS OBTAINED FROM DATA TRAINING WITH A LEARNING RATE OF 0.00002

No	**Data Composition**	**AP**		**mAP**
		Sperm	***Non Sperm***	
1	Train 80%, Validation 10%, Test 10%	86.94%	62.54%	74.74%
2	Train 70%, Validation 25%, Test 5%	87.82%	64.41%	76.11%
3	Train 60%, Validation 20%, Test 20%	88.42%	63.10%	75.76%

Table I, Table II, and Table III shows that the mAP generated based on several data sharing results in an accuracy range of 74% - 79%. The result of the highest training data accuracy is 79.58% which is found in the distribution of 70% test data, 25% validation data, 5% test data, with a learning rate of 0.0002. The results of the lowest training accuracy are 74.74% which are found in the distribution of 80% test data, 10% validation data, 10% test data, with a learning rate of 0.00002. AP results from Sperm and Non Sperm objects differ greatly. This is due to the fact that the number of datasets used for training is imbalanced data. The best AP result in identifying sperm objects is 90.31% and for non-sperm objects it is 68.13%. The biggest results are obtained from the sharing of data and different learning rates.

B. *Graphical Results of Precision, Recall, F1-Score and on Train Model*

The results of the tests carried out in this study obtained the values of precision, recall and F1-score of several types of learning rates. The results obtained in the model training are good, because the results of each precision, recall and F1-score do not experience very large differences in values. Based on each learning rate that produces the highest value, there is a learning rate of 0.0002 with an average value of 0.8 of the precision, recall and F1-score values. The results of precision, recall, and F1-score illustrate that the model that has been made can predict and retrieve information well. The graph can be seen in the Fig. 10, Fig. 11, and Fig. 12.

Fig. 10 shows a graph of the values of precision, recall, and F1-score of three scenarios based on a learning rate of 0.002. It can be seen that the values of precision, recall, and F1-score in the second and third scenarios have a slight difference. However, the second scenario has a value that is superior to precision and F1 scores with values of 0.75 and 0.81. While the third scenario outperformed the recall value of 0.88, only 0.01 difference from the recall in the second scenario.

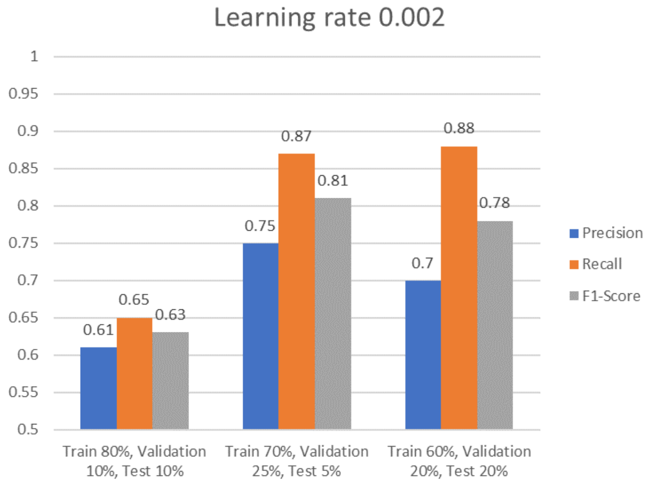


Fig. 10. Graph of Precision, Recall and F1-Score Values at a Learning Rate of 0.002.

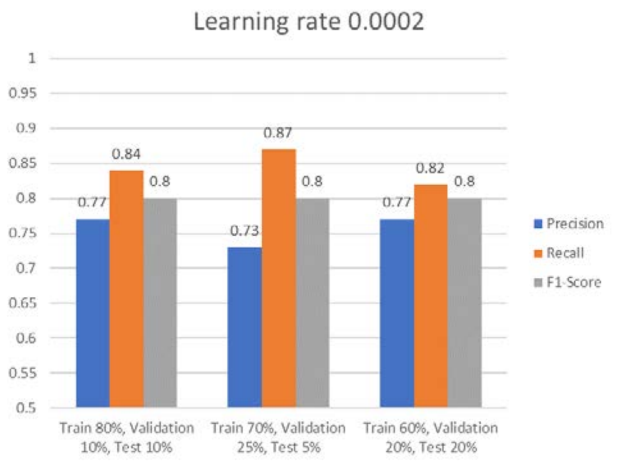


Fig. 11. Graph of Precision, Recall and F1-Score Values at a Learning Rate of 0.0002.

Fig. 11 shows a graph of the values of precision, recall, and F1-score of three scenarios based on a learning rate of 0.0002. It can be seen that the values of precision, recall, and F1-score get a slight difference in the values of the three scenarios. The highest precision value is found in the first and third scenarios with a value of 0.77. then the largest recall value is in the second scenario with a value of 0.87. While the relative F1-Score has the same value in the three scenarios with a value of 0.8.

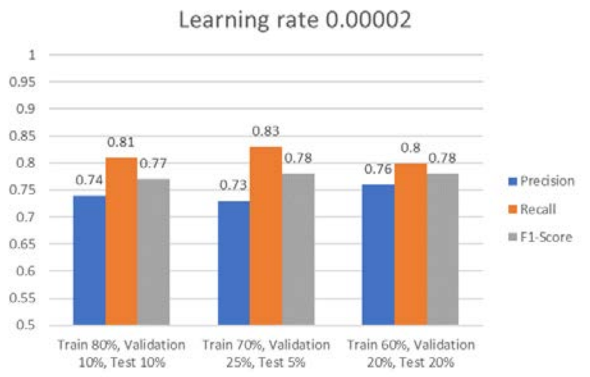


Fig. 12. Graph of Precision, Recall and F1-Score Values at a Learning Rate of 0.00002.

Fig. 12 shows a graph of the values of precision, recall, and F1-scores from three scenarios based on a learning rate of 0.00002. It can be seen that the value of precision, recall, and F1-score in the second scenario has a superior recall value with a value of 0.83. Meanwhile, the highest precision value is found in the third scenario with a value of 0.76. Then for the largest F1-Score value in the second and third scenarios with a value of 0.78.

C. *Overfitting Handling on YOLOv4 Model*

Overfitting occurs when the amount of training data used a slight variation. Based on the training data in this study using training data that has many variations. Training data is taken from several pieces of video frames on the VISEM dataset. YOLOv4 model in doing overfitting analysis according to Alexey the creator of the YOLOv4 model, validation loss in YOLOv4 should not be too much attention because it will tend to decrease continuously, so only mAP accuracy must be considered. If there is a stagnation in the data training process, it is likely that overfitting has occurred. Fig. 13 shows the results of the data training in this research.

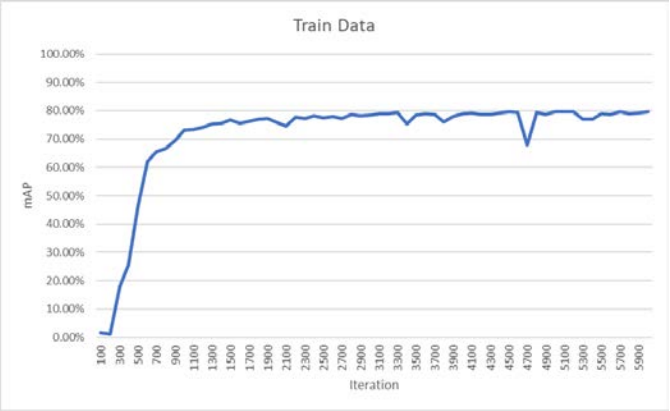


Fig. 13. Graph of Train Data Result from Split Data 70% Train, 25% Validation, and 5% Test with 0.0002 Learning Rate.

Fig. 13 is the result of data training from split data 70% train, 25% validation, and 5% test using a learning rate of 0.0002 which gets the highest mAP results in this research. The graph obtained does not stagnant so that the accuracy of mAP continues to run until the 6000 iteration. The mAP value obtained is 79.58% based on the results of the training data that has been done.

D. *Test Results on the YOLOv4 Model*

The test on this model is done by using a clip from one of the videos from the VISEM dataset. This is because at certain seconds the appearance of the VISEM video dataset will change. This model detects Sperm and Non-Sperm objects in the form of video so that the results of object checking by the model will be displayed based on the frame detected from the video being tested. The following detection results can be seen in Table IV and Table V.

TABLE IV. THE RESULTS OF THE EVALUATION TEST OBJECT ON THE YOLO MODEL

No	Frame Id	Model Prediction		TP	FP	FN
		Sperm	Non Sperm			
1	1	25	8	31	2	0
2	14	25	9	31	3	0
3	26	18	14	30	1	1
4	38	21	13	33	1	0
5	43	23	12	34	1	0
6	55	17	10	26	1	1
7	67	18	9	26	1	0
8	79	19	9	28	0	0
9	86	19	9	27	1	0
10	97	20	9	28	1	0

Table IV shows the result of an evaluation using confusion matrix to get the number of objects that have been detected or not. Detection is collected by object, the results obtained that almost all objects can be detected accurately. The layer capture taken from the YOLO model experiment in the form of video is taken as much as 10 frames, because there are so many frames generated from 1 video.

Table V describes the results of data processing obtained in the process in Table IV. These values are used to obtain precision, recall, AP, and mAP values. The results obtained will be a benchmark for the accuracy of the model in detecting objects. We can see the results of the experiment stated that the model can detect more than 80% of the number of objects in a video frame.

Based on the description in Table IV and Table V, it can be stated that the model can detect sperm and non-sperm objects with good results. In the sperm and non-sperm sections, a bounding box has been successfully created with a video quality of 50 fps, however there are one or two objects that are not legible, this is due to the object being cut off by the video frame. Fig. 14 is a display of the detection results from the model that has been created.

TABLE V. CALCULATION RESULTS OF PRECISION, RECALL, AP, AND MAP

No	Precision	Recall	AP		mAP
			Sperm	Non Sperm	
1	0.94	1.00	0.94	0.73	0.83
2	0.91	1.00	0.99	0.82	0.90
3	0.97	0.97	0.99	0.96	0.98
4	0.97	1.00	0.98	1.00	0.99
5	0.97	1.00	0.99	1.00	0.99
6	0.96	0.96	0.99	1.00	0.99
7	0.96	1.00	0.99	1.00	0.99
8	1.00	1.00	1.00	1.00	1.00
9	0.96	1.00	0.99	1.00	0.99
10	0.97	1.00	0.99	1.00	0.99

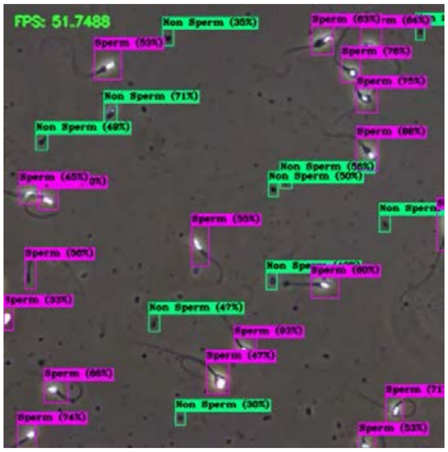


Fig. 14. Display Testing.

IV. CONCLUSION

This research develops an object detection development using the YOLOv4 algorithm to detect sperm and non-sperm objects. Using a dataset derived from the open source Simula Open Dataset, then the data in the form of videos is extracted into 1330 images. In this study, the training process was carried out with 3 different learning rate experiments, namely 0.002, 0.0002, 0.00002. In each of these experiments, 3 data divisions were made for each of the reading rates being tested. The best accuracy results are found in experiments with a learning rate of 0.0002 which has an accuracy value of 79.58% mAP on 70% train data distribution, 25% validation and 5% test. Each trial process for training uses 6000 iterations to create the training data. The test in this study uses video, the results of which are that all objects can be detected properly and have been labeled with a bounding box. In this study there were cases where the model was not able to detect optimally because the video data used contained blurred objects and sperm objects that were cut off by the frame.

ACKNOWLEDGMENT

The experiment in this research used NVIDIA Tesla K80 and Tesla K20 provided by Department of Computer Science, University of Lampung.

REFERENCES

[1] Gatimel, N., Moreau, J., Parinaud, J., & Leandri, R. (2017). Sperm morphology: assessment, pathophysiology, clinical relevance, and state of the art in 2017. ANDROLOGY, 845-862.

[2] Hinting, A., & Agustinus, A. (2021). Technology Updates in Male Infertility Management. Indonesian Andrology and Biomedical Journa, 63-67.

[3] Dhyani, I. A., Kurniawan, Y., & Negara, M. O. (2020). Hubungan Antara Faktor-Faktor Penyebab Infertilitas Terhadap Tingkat Keberhasilan IVF-ICSIi di RSIA Puri Bunda Denpasar Pada Tahun 2017. JURNAL MEDIKA UDAYANA, 2-5.

[4] S.Ningsih, Y. J., & Farich, A. (2016). Determinan Kejadian Infertilitas Pria di Kabupaten Tulang Bawang. Jurnal Kesehatan, 8-5.

[5] Iqbal, I., Mustafa, G., & Ma, J. (2020). Deep Learning-Based Morphological Classification of Human Sperm Heads. Diagnostics, 2-5.

[6] Nissen, M. S., Krause, O., Almstrup, K., Kjærulff, S., Nielsen, T. T., & Nielsen, M. (2017). Convolutional neural networks for segmentation and object detection of human semen. Cornell University, 1-6.

[7] Haugen, T. B., Andersen, J. M., Witczak, O., Hammer, H. L., Hicks, S. A., Borgli, R. J., . . . Riegler, M. A. (2019). VISEM: A Multimodal Video Dataset of Human Spermatozoa. MMSys '19 (ACM SIGMM Conference on Multimedia Systems).

[8] Hicks, t., Andersen, J., Witczak, O., Thambawita, V., Halvorsen, P., Hammer, H., . . . Riegler, M. (2019). Machine Learning-Based Analysis of Sperm Videos and Participant Data for Male Fertility Prediction. Springer Nature, 1-5.

[9] Valiuškait˙e, V., Raudonis, V., Maskeli ¯unas, R., amaševiˇcius, R., & Krilaviˇcius, T. (2020). Deep Learning Based Evaluation of Spermatozoid Motility for Artificial Insemination. Sensors, 2-8.

[10] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. IEEE Xplore, 1-9.

[11] Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. Cornell University, 1-17.

[12] Santoso, A., & Ariyanto, G. (2018). Implementasi Deep Learning Berbasis Keras Untuk Pengenalan Wajah. Jurnal Teknik Elektro, 18, 15.

[13] Rahim, A., Kusrini, & Luthfi, E. T. (2020). CONVOLUTIONAL NEURAL NETWORK UNTUK KALASIFIKASI PENGGUNAAN MASKER. Jurnal Teknologi Informasi dan Komunikasi, 10, 110.

[14] Rohim, A., Sari, Y. A., & Tibyani. (2019). Convolution Neural Network (CNN) Untuk Pengklasifikasian Citra Makanan Tradisional. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 3, 7038.

[15] Alwanda, M. R., Ramadhan, R. P., & Alamsyah, D. (2020). Implementasi Metode Convolutional Neural Network Menggunakan Arsitektur LeNet-5 untuk Pengenalan Doodle. Jurnal Algoritme , 45-56.

[16] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Chen, T. (2017). Recent Advances in Convolutional Neural Networks. ELSEVIER, 8.

[17] Artyani, I. (2019). Simulasi Metode Convolutional Neural Network dan Long Short-Term Memory untuk Generate Image Captioning Pada Gambar Lalu Lintas Kendaraan Berbahasa Indonesia. Journal Teknik Informatika UINJKT, 27.

[18] Peryanto, A., Yudhana, A., & Umar, R. (2020). Klasifikasi Citra Menggunakan Convolutional Neural Network dan K Fold Cross Validation. Journal of Applied Informatics and Computing (JAIC), 45-51.

[19] Rahma, L., Syaputra, H., Mirza, A., & Purnamasari , S. D. (2021). Objek Deteksi Makanan Khas Palembang Menggunakan Algoritma YOLO (You Only Look Once). Jurnal Nasional Ilmu Komputer, 214-217.

[20] Arini, Wardhani, L. K., & Octaviano, D. (2020). Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden. KILAT, 103-114.

[21] Hartanti, D., Kusrini , & Taufiq , E. L. (2018). Penerapan Naïve Bayes Dalams Prediksi Ketercapaian Nilai Kriteria Ketuntasan Minimal Siswa. Jusikom PrimA (Junal Sistem Informasi Ilmu Komputer Prima).

[22] Kusuma, T. A., Usman, K., & Saidah, S. (2021). PEOPLE COUNTING FOR PUBLIC TRANSPORTATIONS USING YOU ONLY LOOK ONCE METHOD. Jurnal Teknik Informatika (JUTIF), 2, 60-64.

[23] Fauziah, D. A., Maududie, A., & Nuritha, I. (2018). Klasifikasi Berita Politik Menggunakan Algoritma K-nearst Neighbor. BERKALA SAINSTEK, 106-114.

[24] Fandisyah, A. F., Iriawan, N., & Winahju, W. S. (2021). Deteksi Kapal di Laut Indonesia Menggunakan YOLOv3. JURNAL SAINS DAN SENI ITS, 10, D26-D30.

Brain Tumor Segmentation and Classification from MRI Images using Improved FLICM Segmentation and SCA Weight Optimized Wavelet-ELM Model

Debendra Kumar Sahoo[1]
Ph.D Scholar, Dept. of ECE
Centurion University of Technology and Management, Bhubaneswar
Odisha, India

Satyasis Mishra[2]*
Dept. of ECE
Centurion University of Technology and Management, Bhubaneswar
Odisha, India

Mihir Narayan Mohanty[3]
Dept. of ECE
SOA University
Odisha, India

***Abstract*—Image segmentation is an essential technique of brain tumor MRI image processing for automated diagnosis of an image by partitioning it into distinct regions referred to as a set of pixels. The classification of the tumor affected and non-tumor becomes an arduous task for radiologists. This paper presents a novel image enhancement based on the SCA (Sine Cosine Algorithm) optimization technique for the improvement of image quality. The improved FLICM (Fuzzy Local Information C Means) segmentation technique is proposed to detect the affected regions of brain tumor from the MRI brain tumor images and reduction of noise from the MRI images by introducing a fuzzy factor to the objective function. The SCA weight-optimized Wavelet-Extreme Learning Machine (SCA-WELM) model is also proposed for the classification of benign tumors and malignant tumors from MRI brain images. In the first instance, the enhanced images are undergone improved FLICM Segmentation. In the second phase, the segmented images are utilized for feature extraction. The GLCM feature extraction technique is considered for feature extraction. The extracted features are aligned as input to the SCA-WELM model for the classification of benign and malignant tumors. The following dataset (Dataset-255) is considered for evaluating the proposed classification approach. An accuracy of 99.12% is achieved by the improved FLICM segmentation technique. The classification performance of the SCA-WELM is measured by sensitivity, specificity, accuracy, and computational time and achieved 0.98, 0.99, 99.21%, and 97.2576 seconds respectively. The comparison results of SVM (Support Vector Machine), ELM, SCA-ELM, and proposed SCA-WELM models are presented to show the robustness of the proposed SCA-WELM classification model.**

Keywords*—*Sine cosine algorithm; extreme learning machine; fuzzy c means; GLCM feature; support vector machine

I. Introduction

Brain tumor-related deaths are increasing worldwide according to the reports of the WHO (World Health Organization). The people affected by brain tumors are suffered from the symptoms of headache, vomiting, mildness of eye vision, and many more as per the medical study. The early treatment of tumor-related disease is essential to avoid recurrent deaths. To make treatment faster, automated segmentation and classification techniques are the requirements for medical diagnosis. The pre-processing step is simple and essential in brain-image analysis. Pre-processing is generally used to reduce the noise and enhance the image resolution and contrast. Many pre-processing approaches are used, like un-sharp masking, veneer filters as well as median-filters. Median filters are usually utilized during the pre-processing phase to protect the boundaries of an image [1]. The image segmentation of brain tumors from "magnetic resonance imaging (MRI)" is a significant assignment for the medical diagnosis of brain tumors. The conventional fuzzy c-means clustering (FCM) algorithm is sensitive to noise. This paper proposes an improved fuzzy local information-based FCM image segmentation to address the difficulties of segmentation. Sehgal et al. [2] proposed a segmentation strategy based on neural network optimization that uses neighborhood attraction using MRI. By taking into account the local attractiveness, the strategy changed the classic FCM method. The enhanced FCM clustering (IFCM), takes into account local attractions, based on two components: characteristics as well as a span of attraction. To partition brain tumor MRIs, researchers used the method published by Nabizadeh et al. [3] for calculating characteristics from the association between the tumors and with brain's LaVs. The method is divided into 4 steps: pre-processing, segmentation, and feature extraction with classification. Li et al. [4] suggested a brain tumor partition technique that incorporated anisotropic diffusion filtering as a pretreatment step, followed by partition as well as tumor extraction utilizing region with circularity using the FCM technique.

Machine learning has ignited considerable interest in modern computers in the field of medicine. In the area of brain-tumor recognition, a variety of advanced machine learning approaches are applied. Advanced methods are employed to identify the use of brain pictures and improve the quality of the information collected, such as image labeling, image reconstruction, skull removal, and registration [13]. As a result, machine learning has enabled clinics, engineers, and computer scientists to collaborate to develop semi-automated and eventually completely automated tumor diagnostic systems with improved accuracy and processing speed. Motivated by

*Corresponding Author.

the advancements of machine learning, we have proposed the following contributions.

A. Contribution of the Research Work

Three contributions are proposed based on image enhancement, image segmentation, and classification. The contributions are summarized as follows:

- In the first aspect, the position and velocity parameters of the SCA algorithm are modified to enhance the quality of the images.
- In the second aspect, the fuzzy factor in FLICM segmentation is replaced with a new fuzzy factor to improve the tumor detection and noise reduction capability from the brain MRI images. The mathematical analysis for the improved fast and robust FLICM segmentation algorithm is presented to authenticate the proposed segmentation.
- In the third aspect, the weights of the Mexican Hat Wavelet -ELM model optimization by the SCA optimization technique are proposed to enhance the classification performance.

The paper is organized as follows. Section 2 presents the research implementation diagram, Section 3 presents the Sine cosine algorithm for image enhancement, the proposed fast and robust FLICM segmentation technique and proposed SCA-WELM model explanation, and Section 4 presents results and discussion of the proposed image enhancement, segmentation, and classification, Section 5 presents the conclusion and followed by the references.

II. Related Work

Several segmentation techniques are presented by the researchers, and some of the latest research is included in the related work. Pinheiro et al. [5] devised a novel MRI technique for detecting brain tumors. Global threshold partitioning has been applied after pre-processing of input MRIs. Before watershed partition, Morphological approaches have been used to improve its results. Elazab et al. [6] proposed an "adaptively regularized kernel-based fuzzy C-means (ARKFCM)" segmentation to reduce computational time than the KFCM segmentation. FCM clustering algorithms with spatial constraints are proposed to remove the noise proposed in [7,8]. Chao et al. [9] proposed a GM-ARKFCM algorithm to show better segmentation than ARKFCM. Cherfa et al. [10] proposed AKFRFCM using Particle swarm optimization to improve segmentation capability. Tao Lei et al [11] presented a "fast and robust FCM (FRFCM)", which uses more parameters, and fails drastically to reduce Gaussian noise, beyond 30%. Satyasis et al. [12] proposed an improved fast and robust FCM algorithm (IFRFCM) to improve the noise reduction capability. Belean et al. [14] proposed a density-based spatial clustering procedure driven by a level-set approach for microarray spot segmentation and quality measures were obtained. Wenxiu et al. [15] proposed a two-phase selective segmentation method, in which the first phase reduces noise on segmentation and the second phase shows the selective segmentation on the preprocessed image.

There are several machine learning methodologies and methods to detect brain tumors by utilizing MRIs. An approach has been described by El-Dahshan et al. [16], during the discovery phase, an artificial feedback neural network and KNN are used. Saritha et al. [17] explained a new method for identifying usual and unusual brain MRI images pathologically. Three features are extracted using wavelet entropy-based spider-web plots. Yang et al. [18] suggested a recent method for MRI-based early recognition of brain tumors. With the RBF function, a kernel-type SVM is used as a classifier. In [19], Kalbkhani et al. employed 2D DWT and generalized autoregressive conditional heteroskedasticity (GARCH). The features are extracted using linear discriminate analysis (LDA), and the feature vectors are reduced using PCA. For the detection process, KNN, as well as SVM identifiers, are used. A method for tumor detection was proposed by Xiao [20] and Abd-Ellah et al. [21]. The input image is used to extract three kinds of characteristics: intensity-based, texture-based, and symmetry-based features. Mohsen et al. [22] employed a recent MRI method for recognizing brain tumors. For dimensionality reduction, PCA is utilized to reduce picture features. A BPNN determines if a subject's pictures are normal or irregular. Soltaninejad et al. [23] suggested a mixed method for MRI recognition of brain tumors. A feedback pulse-coupled neural network is used to preprocess the image (FPCNN). For feature extraction and reduction, PCA and the discrete wavelets transform (DWT) are employed. Using two-level DWT decomposition, the LL sub-band data is delivered to the PCA. For the detection stage, PCA has been utilized to choose a vector of seven features. The FFBPNN is then used to identify whether or not the MRI image is usual or unusual. Abdel-Maksoud et al. gave a new approach for detecting brain tumors using MRIs in [24]. A median filter is used to preprocess MRI images before DWT extracts features. PCA has been used to reduce the number of features, whereas RBF and kernel type SVM has been used to recognize them. Tustison et al. [25] extracted features using Daubechies wavelets, which were subsequently processed using PCA to decrease feature vectors.

The usual and unusual MRIs haave then detected utilizing an SVM as well as RBF. Nabizadeh et al. [26] described a hybrid method to recognize brain tumors utilizing MRIs. The suggested approach involves LS-SVM, GLCM, and noise filtering in three levels: preprocessing, and feature extraction with detection. The four features collected are energy, correlation, homogeneity, and contrast. Huang et al. [27] discovered that the brain MRI dataset is studied using MLP, LVQ, RBF, and SOM classifiers in the recognition step. Median, as well as Gaussian filters are utilized during preprocessing state. Boarder could be extracted by utilizing Gaussian thresholding. GLCM has been utilized to extract features, which results in 21 features that are then reduced to 8 by PCA. Mahima et al. [28] proposed a fractional order contour detection PDE (partial differential equation) model with a regularization term for noise reduction and maintaining the regularity of level set function (LSF). A cellular neural network (CNN) model is used to solve the proposed contour detection PDE. Bogdan et al. [29] proposed an edge-based active contour model (ACM) driven by cellular neural networks (CNNs) for the segmentation procedure. Javeria et al.

[30] proposed the inceptionv3model for deep feature extraction, and quantum variational classifier (QVR) with the 2020-BRATS dataset and achieved more than 90% detection accuracy. Muhammad et al. [31] proposed Berkeley's wavelet transformation (BWT) and deep learning classifier and achieved an accuracy of 98.5%. Ramin et al. [32] Proposed Cascade Convolutional Neural Network (C-ConvNet/C-CNN) classifier with the BRAT-2018 dataset and achieved a dice score of 0.9113 for enhancing tumor. Isselmou et al. [33] deep wavelet autoencoder model with 2500 MR brain images of brat dataset and achieved an accuracy of 99.3% and 0.1 loss validation.

III. Methodology

A. Research Flow Diagram

The research work follows the following steps: (i) The Dataset-255 brain tumor images are collected and the SCA technique is applied for image enhancement and segmented by the novel Improved FLICM segmentation techniques. Further (ii) the segmented images undergo GLCM feature extraction; (iii) in the third stage, the extracted features are fed as input to the proposed SCA weight optimized WELM model for the classification of the benign and malignant tumors; (iv) in the fourth stage classification comparison results of the models are presented. The research implementation phase is shown in Fig. 1.

B. The Sine Cosine Algorithm (SCA)

Swagat et al. [34] proposed PSO and APSO algorithms for enhancement of the gray images, but not applied to brain tumor images. The "sine cosine algorithm" (SCA) [35] is an optimization algorithm based on the sine and cosine functions search also not utilized for brain MRI image enhancement.

According to sine cosine algorithm [35] the position equation is updated as

$$X_i^{n+1} = \begin{cases} X_i^n + \alpha_1 \times sin(\alpha_2) \times |\alpha_3 p^{gbest} - X_i^n|, \alpha_4 < 0.5 \\ X_i^n + \alpha_1 \times cos(\alpha_2) \times |\alpha_3 p^{gbest} - X_i^n|, \alpha_4 \geq 0.5 \end{cases} \quad (1)$$

Where $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are the random variables and α_1 is given by.

$$\alpha_1 = a\left(1 - \frac{n}{K}\right) \quad (2)$$

Where “ is the current iteration, K is the maximum number of iterations”. p^{gbest} is the global best position of the pixels in the image.

The X_i^n signifies the “current position” and X_i^{n+1} signifies the “new position”. The α_1 is the next position in the search, α_2 determines direction of movement, α_3 controls the current movement, and the parameter α_4 uniformly switches among the sine and cosine functions. The parameter and α_1 is given by.

$$\alpha_1 = a\left(1 - \frac{n}{K}\right)$$

Where the current iteration is given by n, maximum numbers of iterations are denoted by K and a is a constant. Now considering the population images Xi as corresponded image position ξ_l in the image and applies the SCA algorithm for image enhancement.

$$\xi_i^{n+1} = \begin{cases} \xi_i^n + \alpha_1 \times sin(\alpha_2) \times |\alpha_3 p^{gbest} - \xi_i^n|, \alpha_4 < 0.5 \\ \xi_i^n + \alpha_1 \times cos(\alpha_2) \times |\alpha_3 p^{gbest} - \xi_i^n|, \alpha_4 \geq 0.5 \end{cases} \quad (3)$$

With the equation operation on the image, the new position of the pixel values of the images is calculated for image enhancement. The pseudo code for the algorithm implementation is presented in Table I.

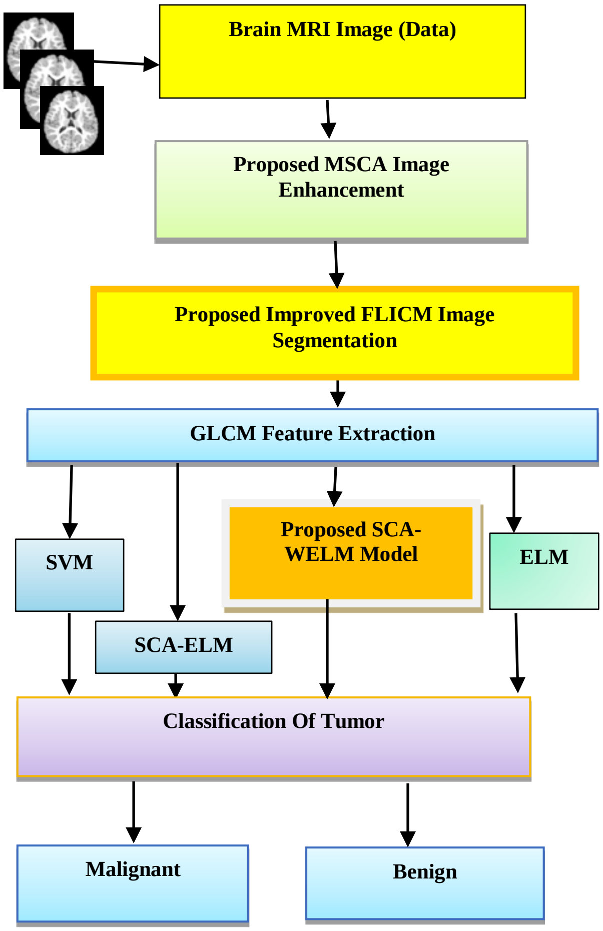


Fig. 1. Research Flow Diagram.

TABLE I. Pseudo Code: SCA Algorithm Implementation for Image Enhancement

Pseudo code: SCA Algorithm implementation for Image
1. Initialize random position and velocity vectors.
2. Initialize the SCA parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4$
3.Evaluate the evaluate fitness based on x_{ii} 4.%optimization loop
5.for i=1:k
6. update SCA parameter to obtain fitness

```
7. update the modified position and velocity equation
$\xi_i^{n+1} = \begin{cases} \xi_i^n + \alpha_1 \times sin(\alpha_2) \times |\alpha_3 p^{gbest} - \xi_i^n|, \alpha_4 < 0.5 \\ \xi_i^n + \alpha_1 \times cos(\alpha_2) \times |\alpha_3 p^{gbest} - \xi_i^n|, \alpha_4 \geq 0.5 \end{cases}$ (4)
8. end for the loop i
9.Stopping criteria: getting fitness as optimal solution
```

C. Improved Fast and Robust Fuzzy Local Information C Means (FRFLICM) Algorithm

According to enhanced fuzzy c means EnFCM [36] algorithm the image ξ is considered from the original image and is given by.

$$\xi_k = \frac{1}{\alpha}\left(x_k + \frac{\alpha}{N_k}\sum_{j\in N_k} x_j\right) \tag{5}$$

Where the gray value of k^{th}pixel of image ξ is given by ξ_k, x_j is neighbors of x_k, N_k is set of neighbors around x_k. Now the new objective function is given by.

$$J_s = \sum_{l=1}^{N}\sum_{k=1}^{c}\gamma_l\, u_{kv}^{m}\|\xi_l - v_k\|^2 \tag{6}$$

Where " u_{il} represents the fuzzy membership of gray value l." γ_l is the number of the pixels having the gray value equal to$l$, and $l = 1,2,\ldots.N$.

According to the FLICM segmentation [37], the fuzzy factor is given by.

$$G_{kv} = \sum_{\substack{k\in N_v \\ v\neq k}} \frac{1}{d_{vk}+1}(1-u_{kv})^m\|x_v - v_k\|^2 \tag{7}$$

To improve the noise reduction capability the fuzzy factor is modified, and the new cost function is given by with improved fuzzy factor as.

$$J_s = \sum_{l=1}^{N}\sum_{k=1}^{c}\gamma_l\, u_{kv}^{m}\|\xi_l - v_k\|^2 + \sum_{v=1}^{N}\sum_{k=1}^{c} {G_{kv}}^2 \tag{8}$$

$$J_s = \sum_{l=1}^{N}\sum_{k=1}^{c}\gamma_l\, u_{kv}^{m}\|\xi_l - v_k\|^2 + \sum_{v=1}^{N}\sum_{k=1}^{c} \left(\sum_{\substack{k\in N_v \\ v\neq k}} \frac{1}{d_{vk}+1}(1-u_{kv})^m\|x_v - v_k\|^2\right)^2 \tag{9}$$

With the new objective function, the segmentation accuracy has been improved and segmentation results are presented in the result section.

D. Wavelet ELM Model with SCA Optimization

In this research work, we propose an SCA optimization for wavelet ELM weight optimization (SCA-WELM) that learns the output weights of the ELM classifier. The pseudo-code for SCA weight optimization of the WELM model is presented in Table II. The SCA-WELM model with hidden and output layers is shown in Fig. 2.

The output function of ELM [38] with L hidden neurons is represented by.

$$y = \sum_{k=0}^{L}\beta_k h_k(w_k; x) \tag{10}$$

where $h(w;x) = [1, h_1(w_1;x), \ldots\ldots., h_L(w_L;x)]$ is the hidden feature mapping and β is the weight vector of all hidden neurons to an output neuron, $h_k(\cdot)$is the activation function of hidden layer. Equation (10) can be written as.

$$H\beta = y \tag{11}$$

Where H is the $N \times (L+1)$hidden layer feature-mapping matrix, whose elements are as follows:

$$H = \begin{bmatrix} 1 & h_1(w_1;x_1) & \cdots & h_L(w_N;x_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & h_1(w_1;x_N) & \cdots & h_L(w_N;x_N) \end{bmatrix} \tag{12}$$

And $h_L(w_N;x_N) = [w_1x_1 + w_1x_1 \ldots\ldots\ldots w_Nx_N].\varphi(t)$

Where $\varphi(t) = c(1-x^2)exp\left(-\frac{x^2}{2}\right)$ and $c = \left(\frac{2}{\sqrt{3}}\pi^{-1/4}\right)$

Equation (10) is a linear system, which is solved by

$$\beta = H^{\dagger}\, d, \qquad H^{\dagger} = (H^TH)^{-1}H^T \tag{13}$$

Where $H^{\dagger}$ is the "Moore–Penrose generalized inverse of matrix H " and $d = [d_1, \ldots\ldots, d_N]^T$.

And $d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_1 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$

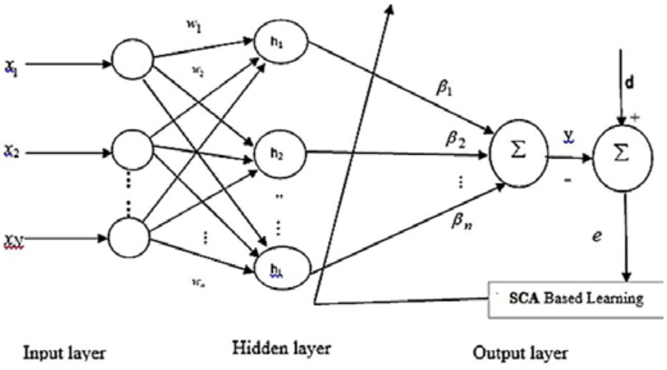


Fig. 2. SCA based WELM Model.

TABLE II. PSEUDO CODE: SCA ALGORITHM FOR WEIGHT OPTIMIZATION OF WELM MODEL

```
PSEUDO CODE:
1. Initialize (ELM weights) randomly.
2.Initialize the SCA position parameters $\alpha_1$,$\alpha_2$, $\alpha_3$
%Starting of the loop
3.Initialize the weights W of the ELM to zero
4.Evaluate the objective function in the next phase to evaluate fitness at first step
5..%Program loop
for i=1:n
for j=1:n
$X_{ij}^{n+1} = \begin{cases} X_{ij}^n + \alpha_1 \times sin(\alpha_2) \times |\alpha_3 y_i^n - X_i^n|, \alpha_4 < 0.5 \\ X_{ij}^n + \alpha_1 \times cos(\alpha_2) \times |\alpha_3 y_i^n - X_i^n|, \alpha_4 \geq 0.5 \end{cases}$
 6. Update $X_{ij}^{n+1}$ for best fit
%Update the weights by using the equation
For j=1:n
$H = \begin{bmatrix} 1 & h_1(w_1;x_1) & \cdots & h_L(w_N;x_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & h_1(w_1;x_N) & \cdots & h_L(w_N;x_N) \end{bmatrix}$
$\beta = H^{\dagger}\, d,$
$H^{\dagger} = (H^TH)^{-1}H^T$
```

7.end for the loop j
8.end for the loop i

9. *Continue till converges, else go to step 4, and repeat until convergence is satisfied.*

IV. Results and Discussion

A. Database Description

The Dataset-255 is collected from "Harvard medical school of architecture (URL: http://med.harvard.edu/ AANLIB/)"[12, 39] which consists of "255 (35 normal and 220 abnormal) 256x256 axial plane brain images" are shown in Table III. "Abnormal brain MR images of Dataset-255 are from 11 types of diseases including Alzheimer’s disease. The Dataset-255 consists of abnormal images of 4 new types of diseases such as chronic subdural hematoma, cerebral toxoplasmosis, herpes encephalitis, and multiple sclerosis".

TABLE III. Details of Dataset-255 [12]

Dataset	Total number of images		Training Images		Testing Images	
	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
Dataset-255	35	220	28	176	7	44

B. Feature Extraction

The "gray-level co-occurrence matrix (GLCM)" [40] statistical features such as "standard deviation, DM (Directional Moment), entropy, coarseness, energy, kurtosis, homogeneity, and energy" features are considered for this research work and presented in Table IV.

TABLE IV. Normalized Feature for Dataset-255

Sl.No.	Features	Values
1	Standard Deviation	0.4587
2	DM	0.8751
3	Entropy	0.9953
4	Coarseness	0.6895
5	Energy	0.7854
6	Kurtosis	0.4122
7	Homogeneity	0.2527

C. Enhancement Results

For the enhancement of the images, the sine cosine algorithm (SCA) has been proposed and compared with the PSO and APSO techniques [34, 41].

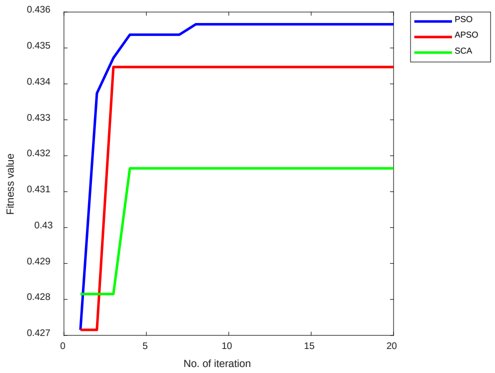


Fig. 3. Image Enhancement of the Benign Tumor Image-1 using PSO, APSO and SCA.

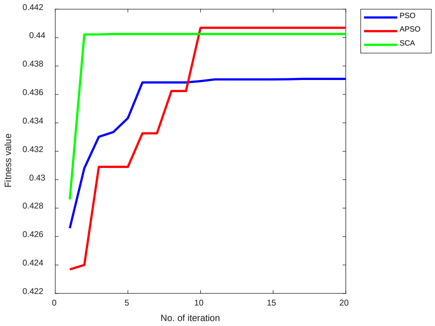


Fig. 4. Image Enhancement of the Benign Tumor Image-2 using PSO, APSO and SCA.

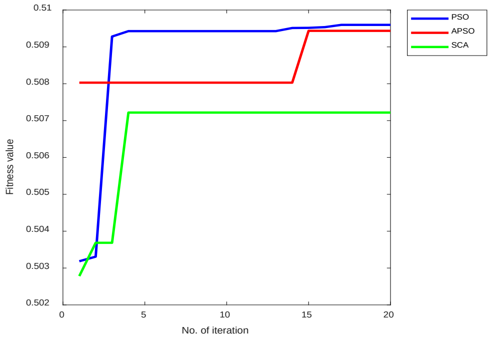


Fig. 5. Image Enhancement of the Malignant-tumor Image-1 using PSO, APSO and SCA.

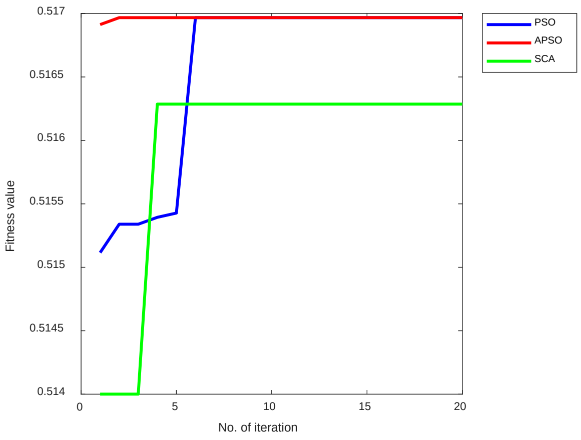


Fig. 6. Image Enhancement of the Malignant Tumor Image-2 using PSO, APSO and SCA.

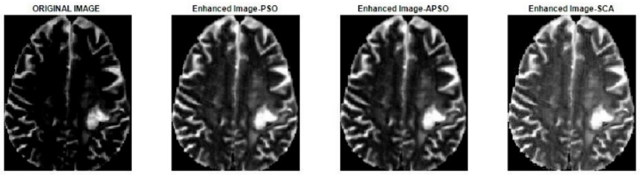


Fig. 7. Benign Tumor Image-1 Image Enhancement using PSO, APSO and SCA.

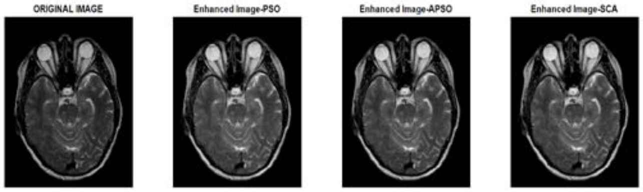


Fig. 8. Benign Tumor Image-2 Image Enhancement using PSO, APSO and SCA.

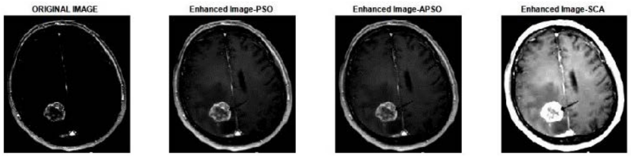


Fig. 9. Malignant Tumor Image-1 Image Enhancement using PSO, APSO and SCA.

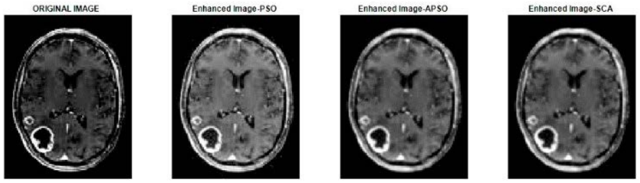


Fig. 10. Malignant Tumor Image-2 image Enhancement using PSO, APSO and SCA.

From Fig. 3 and Fig. 4, it is observed that the fitness value for SCA is 0.4318 and 0.4412 for Benign tumor image1 and image-2, which indicates better image enhancement than the other PSO and APSO methods. Moreover, from Fig. 5 and Fig. 6, it is found that the fitness value is 0.5072 and 0.5162 for malignant tumor image1 and image-2. The higher fitness values for malignant tumor images and lower values for benign tumors indicate better image enhancement of the image. Fig. 7 and Fig. 8 show the benign tumor image-1 and image -2 image enhancement using PSO, APSO, and SCA having fitness values of 0.4318 and 0.4417, and Fig. 9 and Fig. 10 show the malignant tumor image-1 and image -2 image enhancement using PSO, APSO, and SCA having fitness values 0.5072 and 0.5162. Table V presents the fitness values of Benign and Malignant tumor image enhancement.

TABLE V. FITNESS VALUES OF BENIGN AND MALIGNANT TUMOR IMAGE ENHANCEMENT

	Fitness Value	
Algorithm	Benign	Malignant
PSO	0.4367	0.5095
APSO	0.4412	0.5085
SCA	0.4417	0.5162

D. Segmentation Results

The segmentation results are presented in Fig. 11 to Fig. 14. Fig. 11 shows the segmentation of the brain tumor using the FLICM Algorithm. It is visually observed that the noise is not reduced up to the requirement as compared to the other segmentation methods due to the fuzzy factor involvement. Fig. 12 presents the segmentation using the KWFLICM segmentation technique and the segmentation accuracy is 98.48% due to the spatial factor. Fig. 13 shows the segmentation by using the FRFCM Algorithm, which shows a better improvement in terms of accuracy to 98.84% due to the medial filtering in the fuzzy partition matrix and Fig. 14 shows the segmentation by utilizing the proposed improved FLICM technique which has higher accuracy of 99.12% due to improvement in the fuzzy factor. It is visually observed clearly that the proposed Improved FLICM technique has more noise reduction capability than the other segmentation algorithms. The segmentation accuracies are presented in Table VI.

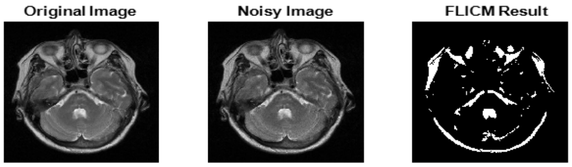


Fig. 11. Segmentation using FLICM Algorithm.

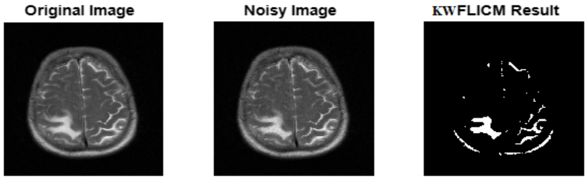


Fig. 12. Segmentation using KWFLICM Algorithm.

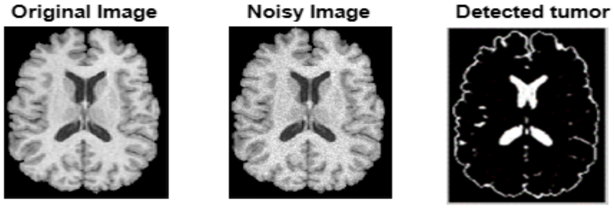


Fig. 13. Segmentation using FRFCM Algorithm.

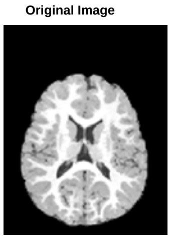


Fig. 14. Segmentation using Proposed Improved FLICM Algorithm.

TABLE VI. IMAGE SEGMENTATION ACCURACY

	Noise level
Algorithm	Speckle Noise
En FCM	97.72
FLICM	98.11
KWFLICM	98.48
NDFCM	98.78
FRFCM	**98.84**
Improved FLICM	**99.12**

E. Quality Measures

To achieve the performance comparison of segmentation, two quality measures are considered "Structural Similarity (SSIM) index and the Quality Index based on Local Variance (QILV) [12]. SSIM is sensitive to the noise and the QILV" is related to the blurring of the edges of the images. On the above, the PSNR (Peak Signal to noise ratio) is also an important parameter related to noise reduction capability. It is observed that the PSNR is 35.39 dB for the proposed FRFLICM segmentation technique which is higher in comparison to the other FCM-based segmentation techniques, which are shown in Table VII. The higher value of PSNR shows a better noise reduction capability. Moreover, the higher value of SSIM and the lower value of QILV shoe the better the segmentation performance.

TABLE VII. QUALITY MEASURES FOR THE MR IMAGE WITH SPECKLE NOISE

	Speckle Noise		
Algorithm	**SSIM**	**QILV**	**PSNR(dB)**
En FCM	0.7748	0.7458	18.14
FLIFCM	0.7894	0.8248	22.14
KWFLICM	0.8287	0.8589	24.52
NDFCM	0.8578	0.8785	26.35
FRFCM	**0.9101**	0.9428	**31.33**
FRFLICM	**0.9758**	0.9541	**35.69**

F. Classifier Performance Measure

Dataset-255 containing T2-weighted magnetic resonance brain images is considered for this research work. To avoid overfitting we have employed a 5×5 cross-validation procedure. "Sensitivity, specificity, accuracy" are the measure of system performance [12].

$$Sensitivity = \frac{TP}{TP + FN} \quad , \quad Specificity = \frac{TN}{TN + FP} \quad ,$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The "5×5-fold cross-validation for each run of Dataset -255" is presented in Table VIII. Table IX shows the "5×5-fold cross validation" procedure for run-1 of Dataset-255. The calculations are considered for the modified SCA-WELM classifier.

TABLE VIII. 5×5 CROSS VALIDATION FOR EACH FOLD DATASET-255 DURING EACH RUN (SCA-WELM CLASSIFIER)

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Total	Accuracy (%)
Run-1	51	51	50	51	50	253	**99.2156**
Run-2	51	51	50	50	51	253	99.2156
Run-3	51	50	50	51	50	252	98.82
Run-4	50	50	51	51	50	252	98.82
Run-5	51	51	51	51	51	255	100
Average Accuracy result							**99.21**

TABLE IX. 5×5 CROSS VALIDATION OF RUN-1 FOR DATASET-255 (SCA-WELM CLASSIFIER

Fold	Test instances	TP	FN	TN	FP	Accuracy (%)
Fold -1	51	43	1	7	0	98.039
Fold -2	51	44	0	7	0	100
Fold -3	51	43	1	7	0	98.039
Fold -4	51	44	0	7	0	100
Fold -5	51	43	1	7	0	98.039
	Average Accuracy result					**99.21**

TABLE X. PERFORMANCE MEASURE OF DIFFERENT CLASSIFIERS

	Dataset -255			
Classifier	Sensitivity	Specificity	Accuracy in (%)	Computational Time in Seconds
SVM	0.96	0.93	96.85	256.2356
ELM	0.97	0.99	97.86	221.3657
SCA-ELM	0.98	0.93	98.97	167.1478
SCA-WELM	**0.98**	**0.99**	**99.21**	**97.2576**

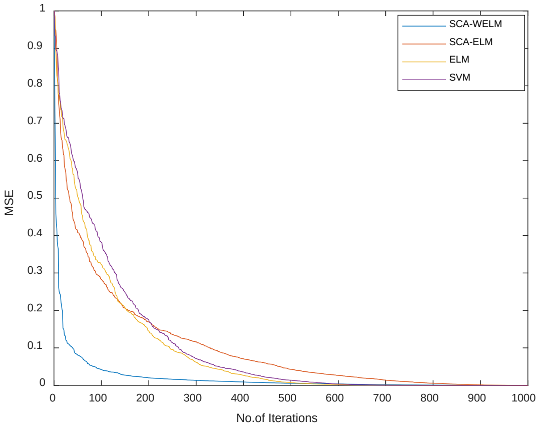


Fig. 15. Mean Square Error Results Comparison.

The proposed SCA-WELM outperforms than other mentioned classifiers in terms of "sensitivity, specificity, and accuracy". The accuracy obtained by SVM, ELM, SCA-ELM, and SCA-WELM is 0.98, 0.99, and 99.21, respectively. Computation time for the proposed SCA-WELM is achieved as 97.2576 seconds (see, Table X). For Dataset-255, it is observed from Fig. 15 that the proposed SCA-WELM model took nearly 360 iterations to converge whereas the SVM, ELM, and SCA-ELM took 530iterations, 580iterations, 430 iterations, and 330 iterations respectively. From the results of mean square error, it is confirmed that the proposed SCA-WELM model provides better performances in terms of accuracy and computational time.

V. Conclusion

In this paper, we have proposed a sine cosine algorithm for image enhancement techniques to improve image quality. The image enhancement technique increases the contrast and smoothens the image by automatic pixel adjustment. A fast and robust FLICM-based segmentation technique has been employed to remove the speckle noise and detect the regions of the brain tumor. The comparison results are presented with other conventional EnFCM, FLICM, KWFLICM, NDFCM, and FRFCM segmentation techniques. The accuracy achieved by the proposed improved FLICM technique shows the robustness of the segmentation technique. Moreover, the higher values of SSIM and PSNR in the case of proposed improved KLICM segmentation confirm the increase in noise reduction capability. The segmented images undergo the GLCM feature extraction technique and the normalized features are presented for classification. The SCA optimization technique has been employed for the optimization of the weights of the wavelet extreme learning machine. The Mexican hat wavelet function is considered in the hidden neurons to increase the capability of classification. Dataset-255 has been considered for this research. The proposed SCA-WELM classifier model has outperformed in classifying the tumors into Benign and Malignant categories. The proposed SCA-WELM model can be applied for breast cancer, and liver tumor medical imaging classification. The proposed model will work for only features as input from the feature extraction methods, but not with images as input to the model like CNN models, which may be a drawback of the research, but this novel method can be applied for different medical images datasets. The novel level set method for 3D brain tumor segmentation [42] and active contour approaches [43] will be the future work of this research to have better visibility and comparison results.

Acknowledgment

We acknowledge Centurion University and Technology and Management, Odisha, India, ECE department staff for availing the laboratory facility and assisted in document preparation.

References

[1] M. Aghalari, A. Aghagolzadeh, and M. Ezoji, “Brain tumor image segmentation via asymmetric/symmetric UNet based on two-pathway-residual blocks,” Biomedical Signal Processing and Control, Vol.69, 2021, 102841. doi:10.1016/j.bspc.2021.102841.

[2] A. Sehgal, S. Goel, P. Mangipudi, A. Mehra, and D. Tyagi. “Automatic brain tumor segmentation and extraction in MR images,” Conference on Advances in Signal Processing (CASP), Vol. 24, pp.104–107, June2016, https://doi.org/10.1109/CASP.2016.7746146.

[3] N Nabizadeh, N John, and C Wright, “Histogram-based gravitational optimization algorithm on single MR modality for automatic brain lesion detection and segmentation,” EXpert Syst. Appl., Vol. 41, pp. 7820–7836.,April 2014, https://doi.org/10.1016/j.eswa.2014.06.043.

[4] Y Li, Q Dou, J Yu, F Jia, J Qin, and PA Heng, “Automatic brain tumor segmentation from MR images via a multi modal sparse coding based probabilistic model. 2015 International Workshop on Pattern Recognition in Neuro Imaging,” Vol. 26, pp.41–49. , June 2015, https://doi.org/10.1109/PRNI.2015.18.

[5] PR Pinheiro, I Tamanini, MCD Pinheiro, and VHC de Albuquerque, “Evaluation of the Alzheimer's disease clinical stages under the optics of hybrid approaches in verbal decision analysis,” Telematics Inform, Vol.35(4), pp. 776–789, Jun 2018, https://doi.org/10.1016/ j.tele.2017.04.008.

[6] A. Elazab, C. Wang, and F. Jia., “Segmentation of brain tissues from magnetic resonance images using adaptively regularized kernel-based fuzzy C-means clustering,” Computational and Mathematical Methods in Medicine, vol. 2015 (5), pp. 485–495, Jul 2015.

[7] H. Xu, C. Ye, and F. Zhang, “A medical image segmentation method with anti-noise and bias-field correction[J],” IEEE Access, Vol. 99, pp. 1, Aug 2020.

[8] A. Kouhi, H. Seyedarabi, and A. Aghagolzadeh, “Robust FCM clustering algorithm with combined spatial constraint and membership matrix local information for brain MRI segmentation,” Expert Systems with Application, Vol. 146, pp. 113159.1–113159.16, May 2020.

[9] Chao Huang, Jihua Wang, "Research on ARKFCM Algorithm Based on Membership Constraint and Bias Field Correction in Neonatal HIE Image Segmentation Method", Mathematical Problems in Engineering, Vol. 2021, pp. 587-596, Aug 2021. https://doi.org/10.1155/2021/4683609.

[10] A. Cherfa, Mokraoui, A. Mekhmoukh and K. Mokrani, "Adaptively Regularized Kernel-Based Fuzzy C-Means Clustering Algorithm Using Particle Swarm Optimization for Medical Image Segmentation," Signal Processing Algorithms, Architectures, Arrangements, and Applications (SPA), pp. 24-29, 2020, doi: 10.23919/SPA50552.2020.9241242.

[11] T Lei, X Jia, Y Zhang, L He, H Meng, and AK Nandi, “Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering,” IEEE Trans Fuzzy Syst, Vol. 26(5), pp. 3027–3041, 2018. https://doi.org/10.1109/tfuzz.2018.2796074.

[12] Satyasis Mishra, Premananda Sahu and Manas Ranjan Senapati, “MASCA- PSO based LLRBFNN Model and Improved fast and robust FCM algorithm for Detection and Classification of Brain Tumor from MR Image” Evolutionary Intelligence, ISSN 1864-5909, Evolutionary Intelligence, Vol.12, pp.647–663 (2020) Springer https://doi.org /10.1007/ s12065-019-00266-x, July,2019,

[13] MH Kayvanrad, AJ McLeod, JS Baxter, CA Mc Kenzie and TM Peters, “Stationary wavelet transform for under-sampled MRI reconstruction. Magn. Reson. Imaging” Vol.32(10), pp:1353–64, 2016. https://doi.org/10.1016/j.mri.2014.08.004.

[14] B. Belean, R. Gutt, C. Costea and O. Balacescu, "Microarray Image Analysis: From Image Processing Methods to Gene Expression Levels Estimation," in IEEE Access, Vol. 8, pp. 159196-159205, Aug 2020, doi: 10.1109/ACCESS.2020.3019844.

[15] Wenxiu Zhao, Weiwei Wang, Xiangchu Feng, and Yu Han, “A new variational method for selective segmentation of medical images,” Signal Processing, Vol. 190, April 2022, ISSN 0165-1684. https:// doi.org/10.1016/ j.sigpro.2021.108292.

[16] ESA El-Dahshan , T. Hosny, and ABM Salem, “Hybrid intelligent techniques for MRI brain images classification,” Digit. Signal Processing, Vol. 20(2), pp. 433–41, 2010. https://doi.org/10.1016/j.dsp.2009.07.002.

[17] M Saritha, KP Joseph, and AT Mathew, “Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural net-work,” Pattern Recogn. Lett.;Vol. 34(16), pp. 2151–6, Jul 2013. https://doi.org/10.1016/j.patrec.2013.08.017.

[18] G Yang, Y Zhang, J Yang, G Ji, Z Dong, and S Wang, “Automated clasification of brain images using wavelet-energy and bio geography-based optimization, Multimed. Tools Appl., Vol. 26, pp. 1–17, May 2015. https://doi.org/10.1007/s11042-015-2649-7.

[19] H Kalbkhani, MG Shayesteh, and B Zali-Vargahan, “Robust algorithm for brain magnetic resonance image(MRI) classification based on GARCH,” V ariances series. Biomed.Signal Process. Control, Vol. 8(6), pp. 909–19, Jun 2013. .https://doi.org/10.1016/j.bspc.2013.09.001.

[20] K Xiao, A Liang, HB Guan, and AE Hassanien, “Extraction and application of de-formation-based feature in medical images,” Neuro computing 2013, Vol. 120 (SupplementC), pp. 177–84. https://doi.org/10.1016/j.neucom.2012.08.054.

[21] Abd-Ellah MK, Awad AI, Khalaf AAM, and Hamed HFA, “Two-phase multi-model automatic brain tumour diagnosis system from magnetic resonance images using convolution neural networks,” EURASIPJ. Image Video Process. 2018, Vol. 97(1), pp. 1–10, May 2018. https://doi.org/10.1186/s13640-018-0332-4.

[22] H Mohsen, ESA El-Dahshan, El-Horbaty ESM, and ABM Salem. “Classification using deep learning neural networks for brain tumors,” Future Comput. Inf. J. 2018; Vol. 3, pp. 68–71, Nov 2018. https://doi.org/10.1016/j.fcij.2017.12.001.

[23] M Soltaninejad, G Yang, T Lambrou, N Allinson, TL Jones, and TR Barrick, “Automated brain tumor detection and segmentation using super piXel-based extremely randomized trees in FLAIRMRI,” Int. J. Comput. Assist. Radiol. Surg., Vol. 12(2), pp. 183–203, Feb 2017. https://doi.org/10.1007/s11548-016-1483-3.

[24] E Abdel-Maksoud, M Elmogy, and R Al-Awadi, “Brain tumor segmentation based on a hybrid clustering technique,” Egypt. Inf. J. 2015; Vol. 16(1)pp. 71–81, 2015. https://doi.org/10.1016 /j.eij.2015 .01.003.

[25] NJ Tustison, KL Shrinidhi, M Wintermark, CR Durst, BM Kandel, and JC Gee, “Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR,” Neuro informatics, Vol. 13(2)pp. 209–25, April 2015. https://doi.org/ 10.1007/s12021-014-9245-2.

[26] N Nabizadeh and M Kubat, “Brain tumors detection and segmentation in MR images: Gabor Waveletvs. Statistical features,” Comput. Electr. Eng.2015; Vol. 45, (Supplement C), pp. 286–301, April 2015. https://doi.org/10.1016/j.compeleceng.2015.02.007.

[27] M Huang, W Yang, Y Wu, J Jiang, W Chen, and Q Feng, “Brain tumor segmentation based on local independent projection-based classification,” IEEE Trans. Biomed. Eng. 2014, Vol. 61(10), pp. 2633–45, 2014. https://doi.org/10.1109/TBME.2014.2325410.

[28] M Lakra and, S. Kumar, “A fractional-order PDE-based contour detection model with CeNN scheme for medical images,” J Real-Time Image Proc Vol. 19, pp. 147–160, 2022. https://doi.org/10.1007/s11554-021-01172-1.

[29] Bogdan Belean, “Active Contours Driven by Cellular Neural Networks for Image Segmentation in Biomedical Applications,” Studies in Informatics and Control, ISSN 1220-1766, Vol. 30(3), pp. 109-120, 2021. https://doi.org/10.24846/v30i3y202110.

[30] J Amin, MA Anjum, M Sharif, S Jabeen, S Kadry and P Moreno Ger, “A New Model for Brain Tumor Detection Using Ensemble Transfer Learning and Quantum Variational Classifier," Comput Intell Neurosci. 2022, Vol. 21, Vol. 42, April 2014. doi: 10.1155/2022/ 3236305. PMID: 35463245; PMCID: PMC9023211.

[31] Muhammad Arif, F. Ajesh, Shermin Shamsudheen, Oana Geman, Diana Izdrui, and Dragos Vicoveanu, "Brain Tumor Detection and Classification by MRI Using Biologically Inspired Orthogonal Wavelet Transform and Deep Learning Techniques", Journal of Healthcare Engineering, Vol. 2022, pp. 18 ,Jun 2022. https://doi.org/10.1155/2022/2693621.

[32] R.Ranjbarzadeh, , A. Bagherian Kasgari, , Jafarzadeh Ghoushchi, “Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images,” Sci Rep 11, 10930 Vol. 31, pp. 568-574, Aug 2021. https://doi.org/10.1038/s41598-021-90428-8.

[33] I.Abd El Kader,; Xu, G.; , Z.; Shuai, S.Saminu,; I.S.I.Javaid,;Ahmad, and S. Kamhi, “Brain Tumor Detection and Classification on MRImages by a Deep WaveletAuto-Encoder Model,” Diagnostics 2021, Vol. 11, pp. 1589, 2021. https://doi.org/ 10.3390/diagnostics11091589.

[34] SK Behera, S. Mishra and D Rana , “Image enhancement using accelerated particle swarm optimization, Int J Eng Res Techno, Vol.. l4, pp. 1049–1055, Jun 2015.

[35] M. Seyedali, “A Sine Cosine Algorithm for Solving Optimization Problems, Knowledge-Based Systems” Vol. 25, pp. 521-524, Aug 2016, doi: 10.1016/j.knosys.2015.12.022.

[36] L Szilagyi, Z Benyo, SM Szilagyii, and Adam HS, “MR brain image segmentation using an enhanced fuzzy c-means algorithm”. In: Proceeding of the 25th annual international conference of the IEEE EMBS, pp 17–21, April 2003.

[37] S Krinidis and V Chatzis “A robust fuzzy local information c-means clustering algorithm”. IEEE Trans Image Process Vol.19(5), pp. 1328–1337, Jul 2010. https://doi.org/10.1109/ tip. 2010.2040763 AQ7.

[38] E Soria-Olivas., J Gomez-Sanchis., J. D. Martin, J. Vila-Frances, M Martinez., J. R Magdalena and A. J Serrano, “BELM: Bayesian extreme learning machine,” IEEE Trans. Neural Netw., vol. 22, (3), pp. 505–509, Mar. 2011.

[39] Dataset: Webpage of Medical School of Harvard University. www.med.harvard.edu/AANLIB/home.html.

[40] DR Nayak, R Dash and B Majhi, “Discrete ripplet-II transform and modified PSO based improved evolutionary extreme learning machine for pathological brain detection,” Neurocomputing, Vol. 28, pp. 288-299, April 2016. https://doi.org/10.1016/j.neucom.2017.12.030.

[41] S.Mishra, J.Gelmecha Demissie, Ram S Singh., Davinder Singh Rathee and T Gopikrishna., Hybrid WCA–SCA and modified FRFCM technique for enhancement and segmentation of brain tumor from magnetic resonance images”, Biomedical Engineering: Applications, Basis and Communications, Vol. 33 (3), Jun 2021. DOI: 10.4015/S1016237221500174.

[42] Chaima Dachraoui, Aymen Mouelhi and Salam Labidi, “Brain MRI monitoring approach of lesion progress in multiple sclerosis using active contours,” International Journal Of Modelling Identification And Control Vol. 38 (1), pp. 32-45, 2019.

[43] A Khosravanian, M.Rahmanimanesh, and P. Keshavarzi, “Level set method for automated 3D brain tumor segmentation using symmetry analysis and kernel induced fuzzy clustering,” Multimedia Tools Appl 81, pp. 21719–21740, April. 2022. https://doi.org/10.1007/s11042-022-12445-7.

A Machine Learning and Multi-Agent Model to Automate Big Data Analytics in Smart Cities

Fouad SASSITE, Malika ADDOU, Fatimazahra BARRAMOU
Team (ASYR) - Laboratory of Systems Engineering
Hassania School of Public Works (EHTP)
Casablanca, Morocco

***Abstract*—The objective of this paper is to present an architecture to improve the process of automating big data analytics using a multi-agent system and machine learning techniques, to support the processing of real time big data streams and to enhance the process of decision-making for urban planning and management. With the rapidly evolving information technologies, and their utilization in many areas such as smart cities, social networks, urban management and planning, massive data streams are generated and need an efficient approach to deal with. The proposition in this paper adopts the concept of smart data which focuses on the value aspect from big data. The proposed architecture is composed of three layers: data acquisition and storage, data management and processing and the service layer, based on a multi-agent system to automate the big data analytics; the proposed model describe the functionalities of the system and the collaboration between agents, these autonomous entities receive data streams in real time, they perform operations of preprocessing, big data analytics and storage into a Hadoop cluster. The techniques of machine learning are also used to enhance the process of decision making, such the use of classification algorithms to predict habitat type based on the characteristics of a population to help making efficient urban planning decisions. The proposed system can serve as a platform to support data management and to conduct effective decision-making in smart cities.**

Keywords—Big data analytics; machine learning; smart data; multi-agent system; automation; decision-making; urban planning

I. Introduction

With the increasingly rapid evolution of information technologies, and its use in many areas such as smart cities, social networks, online business applications, transportation data, urban management, and planning. These massive data are generated with high velocity and require real-time processing, the emergence of the concept of Big Data in many areas has become a reality that imposes itself on systems based on traditional data management and processing technologies such as relational databases. The amounts of generated data are increasingly unstructured due to the diversity and the heterogeneity of the data sources, some studies [1] are claiming that the use of connected object can reach more than 25 billion in 2020. This rapid growth of generated big data imposes new challenges in terms of storage, data analytics and processing, it requires new approaches to provide more reliable solution to deal with the big data.

Useful information, actionable knowledge and valuable data is normally incorporated into these voluminous raw data, Smart Data is the approach that focuses on the value aspect of the big data and try to exploit these huge masses and discover knowledge from these data [2].

Smart data with a focus on veracity and value has been introduced, The goal is to effectively clean or rectify imperfections in the raw data and put the action on the valuable data, which can be effectively used by businesses and governments for planning, evaluation, control and intelligent decision making [3], [4], in order to turn big data into smart data, some researches [5] focuses on the importance of the preprocessing steps and others [6] proposes to highlight the importance of improving four fields of interest:

- Reliable infrastructures,
- Data Organization and management,
- analyze and prediction,
- Decision support and automatization.

This research is in line with this theme of research, and it tries to give a contribution in this area by proposing a multi-agent system in order to automate the processing and analysis of Big Data and to improve the decision-making process, to better use the collected Big Data and to give a sense of these data.

The use of this paradigm in smart cities constitutes an added value in the process of data management and the exploitation of the most of these data to refine the decision making in real time, for problems that require actions in real time and in an automatic way without human intervention based on autonomous agents dotted with artificial intelligence to solve problems related to smart cities such as the congestion of traffic flows, or for the decision making in the long and medium term such as to predict and anticipate the equipment and the necessary infrastructures for the urban planning.

The applications of this proposition can be in several domains notably in the field of urban management and planning, the use of this system can improve the handling of big data issued from the management of smart cities and help to exploit the amount of data generated and analyze it in an intelligent way to create machine learning models that can guide efficient decision making in these cities.

This paper is structured as follows: Section II will discuss some paradigms related to this research topic. Section III will highlight and discuss some related works. Section IV describes the proposed approach. Section V will study the conducted experimentations in this work and the final section will conclude and gives some perspectives.

II. Background of the Study

This section will highlight some of the main concepts related to this research work, First, we will highlight the complexity and diversity of the nature of big data from smart cities, and then show the value of smart data which focuses on the value aspect of the big data and on intelligent processing. Next, we will expose the importance of multi-agent systems in the process of big data processing automation in order to automate the process of real time decision making. Finally, we will explain the importance of machine learning techniques and algorithms, to strengthen the cognitive part of the proposed system.

A. *Big Data to Smart Data*

Big data usually refers to data with large volume of datasets characterized by the complexity and challenges to handle it due the nature of those data and the actual technologies to store and process this type of data [7], [8] the main characteristics in the literature to classify the data as big data:

- Volume, Variety velocity shows the manner in which the data has been generated and the process of storing and processing it.
- Value and Veracity focuses on value aspect, the usefulness, and the quality of data.

In order to automate the process of big data analytics a study of the nature of this data should be done to understand it, basically the big data sources generates three main formats [9] as described in the Fig. 1.

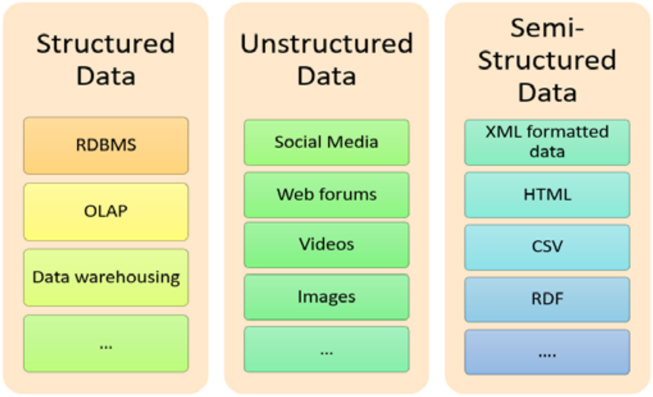


Fig. 1. Taxonomy of Big Data Types.

Smart data (value-based) has been introduced, to highlight valuable data, which can be effectively used by companies and governments for effective planning, monitoring, evaluation, reporting and intelligent decision-making. Three core attributes are necessary for data to be intelligent: it has to be accurate, actionable and agile [10], [11].

A. *Multi-agent System*

Multi-agent System can be defined as a collection of autonomous entities know as agents [12]. The agents can in a collaborative way solve complex problems [13], [14] they can interact with the environment and the other agents to achieve goals or the complete tasks the agents are characterized with:

- Sociability: To reach their goals they can share their knowledge or request useful information other agents.
- Autonomy: The agents can perform some appropriate actions and executing the process of decision-making.
- Proactivity: The agents can perform effective actions by using their historical data, the data from the sensors or from other agents or their environment.

The real value added of agents can be reached through the use of the collaborative work of each agent to solve complex problems.

In this work the framework JADE is used, [15] which is a multi-agent system platform that provides the infrastructure to deploy agents, this platform assures this functionality with the help of other components like:

- Directory Facilitator: DF is the component responsible for providing yellow pages services in the platform, it allows agents to publish their services by sending requests of registration, and this operation allows other agents to collaborate through the published services.
- Agent Management System: AMS is a necessary component in an agent platform; this module is responsible for managing the agents, and assuring the operations of creation, destruction, migration, etc. of the agents in the platform.
- Message Transport Service: this service is responsible of transporting messages between agents in the platform; these messages meet the standards described in the FIPA-ACL.

B. *Machine Learning*

Machine learning is field of Artificial intelligence that relates to a wide range of algorithms for making intelligent predictions based on a data set. These data sets are often characterized with large volume [16], Recent advances in machine learning field have achieved what seems like a human standard of semantic understanding and information extraction and some ability to sense abstract patterns with higher accuracy than human experts [17], [18].

The machine learning techniques are nowadays widely used in different domains like the computer vision, prediction, classification, recommendation, semantic analysis, natural language processing and information retrieval [19] [16], [20].

III. Related Work

This section will highlight some research works with relation of this research question.

In the literature, several researches are done to study the processing and management of big data.

With the technological advancements we are experiencing today, the large amounts of data generated are multiplying rapidly, and the process of manipulating these data is complicated, hence the need to automate the process is justified in several areas like the in industry [21], [22] or in the urban planning and management domain [23].

The use of big data analytics[24] is widely demanded in multiple areas such as security and intrusion detection [25], healthcare [26], The proposition of [27] aims to provide a process composed of the two sub-processes the first for big data management and the second for the big data analytics each sub-process is composed of set of operations described as bellow:

- Big Data Management:
    - Acquisition and recording,
    - Extraction, cleaning, and annotation,
    - Integration, Aggregation and Representation.
- Big data analytics:
    - Modeling and Analysis,
    - Interpretation.

The authors [28], [29] proposed a method based on the generation of workflows of data processing based on a service oriented approach[30], the idea is to divide the totality of the functionalities of the system into a set of services, these services are organized according to four main phases:

- Planning stage,
- Discovery stage,
- Selection stage,
- Execution stage.

The system selects according to the request received a set of services in each phase and then generates a workflow based on the selected services

Other authors [31] proposed a the main steps for extracting value form big data through the definition of four steps:

- Gather data,
- Load data,
- Transform data
- Extract data.

Several works presented in the literature try to give the necessary steps for the system to realize the tasks of analysis and big data processing, but they don't put the action on giving the possibility to the system to have an adaptive behavior according to the nature of the processing and the requests received by the system in an autonomous way, and they need a human intervention to program and manage tasks. The use of the multi-agent paradigm in this sense can enhance the cognitive capabilities of the system of learning efficient ways to process and analyze data and to help in improving the processes of decision-making based on a knowledge base.

IV. Proposed Approach

In order to increase the efficiency of the big data management and analytics process and to implement a new strategy to improve the analysis and processing operations through automation, a model based on three layers architecture and a multi-agent system is proposed.

A. *The Proposed Model*

The proposed model is mainly based on a multi-agent system and adopts the Smart Data approach that focuses on the value aspect of data in order to optimize the exploitation of the large masses of data stored on big data storage clusters. This model has the objective of automating the processing of big data and improving the efficiency of decision making in this context using machine learning models.

As presented in the Fig. 2, this architecture is based on a collection of autonomous entities to constitute a multi-agent system. The multi-agents are organized according to three-layer architecture:

- The data acquisition and storage layer: This layer is responsible for interacting with the heterogeneous data sources that usually a system handling big data can interact with, for example in the field of smart cities management often the system interacts with data coming from different sources whether it is from sensors or external data sources, these data being of various types and generated at different rates, in batches or in real time streams.

This layer also allows to manage the storage of the generated or collected data platform through the use of an HDFS storage cluster which is distributed over several nodes in order to guarantee high availability and also to distribute the data processing operations.

There is also a knowledge base that can help the agents in making decisions and choosing the necessary processing step for each request received by the system.

- The data management and processing layer: In this layer, there are principally the modules dedicated to data processing and analysis, such as components for pre-processing operations like cleaning, filtering, normalization, transformation, and data reduction, as well as the components for data processing like the entities for data processing and analysis in batches specially for historical data and stream processing in the case when the system deals with real time streams of data.
- The data services layer: The main role of this layer is to communicate the results of the processing performed by the system as a result of a query to the applications and services, the monitoring of dashboards and the production of reports, as well as to provide interfaces that facilitate the interaction of the applications with the proposed system.

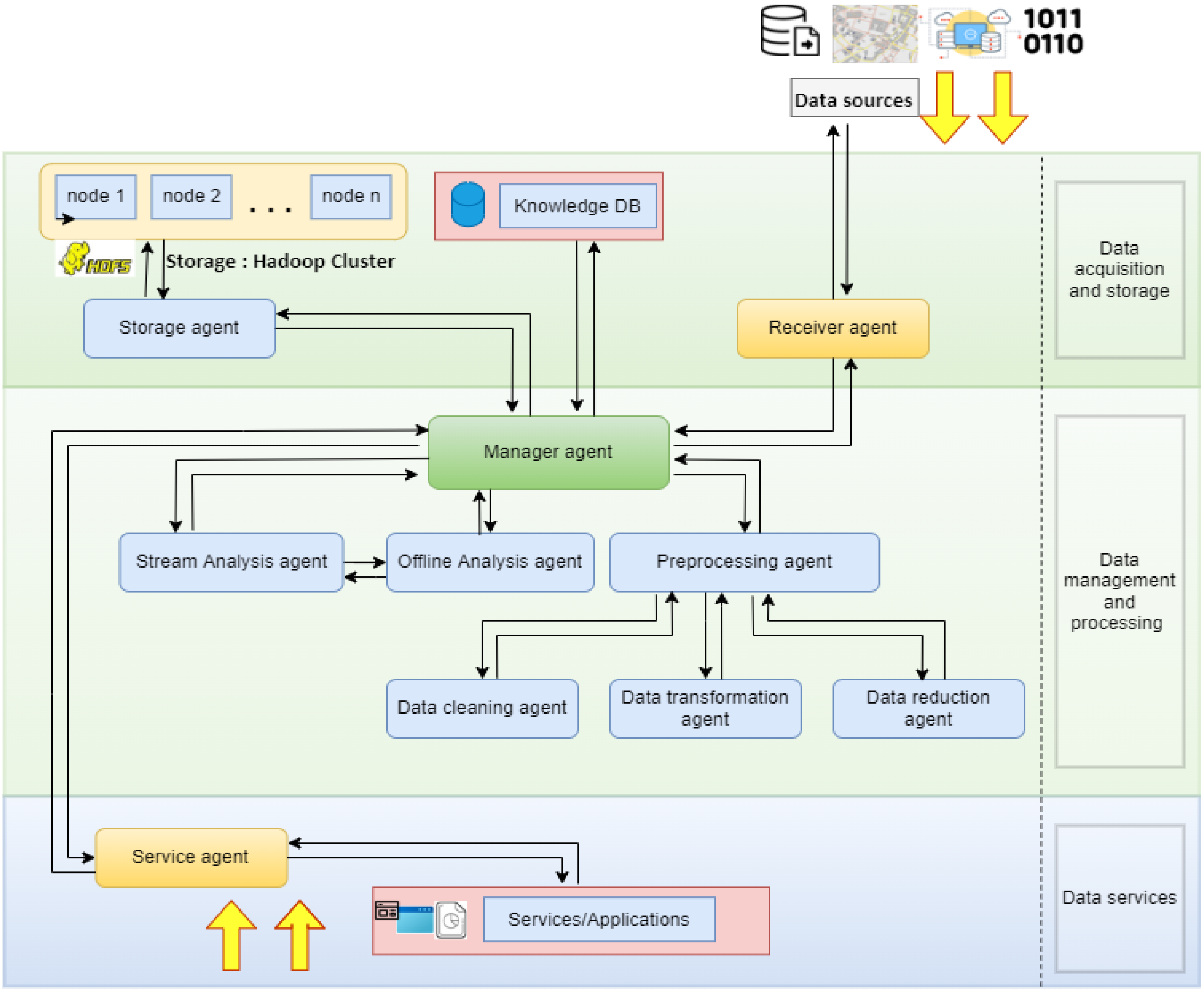


Fig. 2. The Proposed Multi-layer Architecture for the MAS.

Such functionality can help improve and automate the decision-making process, as well as the management and processing of real-time data. The proposed system can serve as a platform to support data management and to conduct effective decision-making in smart cities.

B. Components of the Proposed Multi-agent System

The proposed multi-agent system, through the cooperation of the agents, can automate the processing and the management of the data, in accordance with the three layers of the system: data acquisition, data management and processing, and the services and communication layer.

The system is composed of multiple agents:

- Receiver agent: This agent represents the system's interface with the data sources. The receiving agent handles data of multiple types and from different sources, such as sensors, external databases, web services, etc. and at different rates, streaming or batch.
- Storage Agent: this agent is in charge of all operations related to storage management and data reading from the Hadoop cluster.
- Service Agent: This agent is designed to interact with applications and services and to communicate the processing results through user friendly interfaces and applications. With this agent, the end-users can also schedule or send requests or execute certain tasks.
- Offline Analysis Agent: This agent allows the system to perform operations and tasks related to data batch processing, it can handle the requests of data analysis using voluminous historical data, and usually those data are collected and stored into the cluster storage. Those tasks of processing can help to automate the process of scheduled or per period data analysis tasks.
- Stream Analysis Agent: the main function of this agent is to process data streams that are usually in real time. The use of technologies that are generating data in real time such as sensors and devices that are transmitting requests and data streams in real time. These queries expect responses from the system in real time.

The collaboration between the two agents responsible of processing: the stream Analysis Agent and the Offline Analysis

Agent can resolve complex problems of data processing for example the requests they require to handle data streams in real time and needs to perform operations on the historical data already stored on the system, this collaboration is carried by the Manager Agent.

- Knowledge Base Component: is a database used by the multi-agent system to store rules to facilitate the selection process of the adaptive behavior of the agents based on the type of the request and the data to process.
- HDFS storage Cluster: is the component of the system that supports distributed data storage through the use of many data blocks or partitions. This distributed file system stores the data using several nodes, called data nodes, and managed by the name node which try to balance the load between theses nodes.

The HDFS also assure the replication of the data to guarantee high availability. adopting this type of storage is more appropriate due to requirement of data processing and to the nature of the big data constraints [32], especially when the system deals with fast processing queries with low latency.

- Preprocessing Agent: The preprocessing stage is considered as a primordial phase that help extract to value aspect from the raw big data. The smart Data approach focuses and consider it like a key to prepare the data, since the data generated from real world applications are imperfect and may contains some redundancies, inconsistencies, and noisy values.
- Data Cleaning Agent: This is an agent that checks the received data and imposes a strategy to clean the received dataset. The Smart data aspect focuses on the preprocessing steps and the cleaning phase is an important one with the aim to remove noisy and redundant instances [33] to allow the system to operate automatic learning algorithms on reliable datasets.
- Data Transformation Agent: This agent is designed to change the data format according to the processing needs, to smooth out the training phase by applying data normalization, aggregation, and filtering operations.
- Data Reduction Agent: This agent has an essential contribution in the work of the system by reducing the time of data processing which will impact the positively the process by reducing the time latency, deleting the faulty data and reducing the volume, in this phase there are several techniques in the literature [34] that can be applied with the aim to simplify the datasets dimensionality reduction, discretization and instance reduction.
- Manager Agent: The manager agent has in important role in the proposed system it represents the main component responsible of coordination of the tasks sharing between the other agents.

This entity can take an adoptive behavior based on the nature of the requests received and a knowledge base.

C. Operating Principle

The proposed multi-agent system described previously shows the necessary steps to assure the functioning of the system in different situations.

The system can interact with external applications and data sources and subsequently initiate or execute data analysis operations mainly by triggering two agents, the one responsible for data acquisition or services:

- Receiving a request from a sensor generating a data flow or from an external api or service applies the collaboration of the agents of the system.

Once the request is received by the receiving agent and sends it to the Manager, the latter distributes the tasks to store and process this request. The tasks of pre-processing, online analysis and offline analysis are performed to respond to this request in real time, after that the service agent can transfer the result.

- Receiving a request from the application layer as in the case of scheduling a task or generating a report or a specific request to execute.

The Service agent sends the request to The Manager agent which decides to collect the necessary data from the data cluster or to retrieve it from the data sources; and according to the type of this request, it makes the decision to send this data to one of the data processing agents, either in stream or in offline analysis and returns the result through the service agents.

V. Experimentation and Tests

The proposed approach consists in developing a multi-agent model for the automation of big data processing and analysis by applying the smart data approach.

This section attempts to highlight the adaptive behavior of the proposed system by the study of three cases of use of the system.

The system is composed a set of agents linked to stored algorithms that are defined by the organization's business rules. These agents will be distributed across multiple nodes in order to promote task distribution and load balancing across nodes.

For the implementation of this approach the JADE [15] Framework was adopted.

The communication and the interactions between agents is principally done through the exchange of messages [35] each message is labeled with an act of communication such as: request, inform, notify, propose, etc.

The Fig. 3 represents a snapshot of the behavior of the multi-agent agent system at a time t, where the different communications between the agents of the proposed system can be observed, this communication can be done by sending ACL messages or by sharing information in the environment of the agents, this communication can be in the context of a collaboration, coordination, or negotiation, the agents adopt an adaptive behavior according to the nature of the data and the requests received by the system.

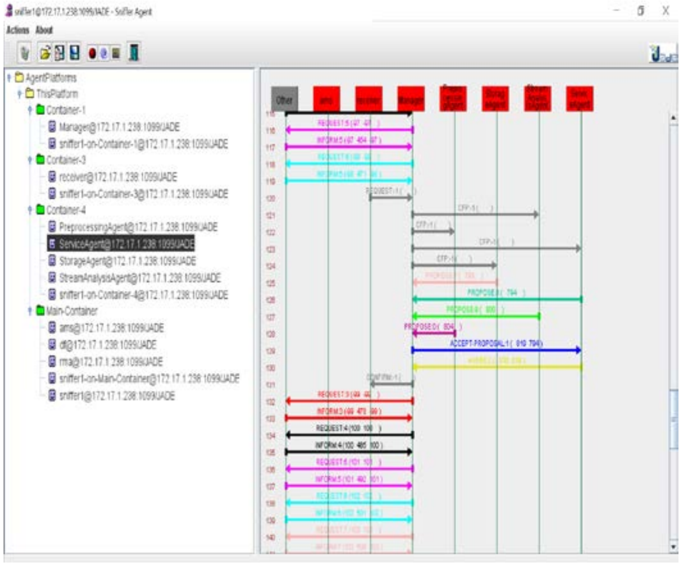

sniffer1@172.17.1.238:1099/JADE - Sniffer Agent
Actions About
AgentPlatforms
ThisPlatform
Container-1
Manager@172.17.1.238:1099/JADE
sniffer1-on-Container-1@172.17.1.238:1099/JADE
Container-3
receiver@172.17.1.238:1099/JADE
sniffer1-on-Container-3@172.17.1.238:1099/JADE
Container-4
PreprocessingAgent@172.17.1.238:1099/JADE
ServiceAgent@172.17.1.238:1099/JADE
StorageAgent@172.17.1.238:1099/JADE
StreamAnalysisAgent@172.17.1.238:1099/JADE
sniffer1-on-Container-4@172.17.1.238:1099/JADE
Main-Container
ams@172.17.1.238:1099/JADE
df@172.17.1.238:1099/JADE
rma@172.17.1.238:1099/JADE
sniffer1-on-Main-Container@172.17.1.238:1099/JADE
sniffer1@172.17.1.238:1099/JADE
Other
ams
receiver
Manager
REQUEST:5
INFORM:5
CFP:1
PROPOSE:4
PROPOSE:0
ACCEPT-PROPOSAL:1
REQUEST:3
INFORM:3
REQUEST:4
INFORM:4
REQUEST:6
INFORM:5

Fig. 3. The Communication between the Agents of the Proposed System.

This part will study the adaptive behavior of the proposed multi-agent system by illustrating three case studies where an explanation of the operating principle of the system and the flow of processing that changes depending on the nature of the request received by the system will be presented, the three scenarios are related to the field of management and planning of smart cities.

The first is to study the behavior of the system using a large dataset that contains data from the last census in Morocco for the city of Casablanca. The used data contain information and characteristics about the habitat and individuals in order to build a machine-learning model to predict the type of habitat according to the evolution of individuals. Thus, predict the equipment and infrastructure necessary for the urban planning of this city.

The second scenario is to show the behavior of the system and its use for listening to changes and storing data from different data sources.

And the third scenario is to use the system for real-time prediction tasks using the machine learning model already created in the first scenario.

A. Case 1: Training a Machine Learning Model from a DataSet Stored in HDFS

In this case the request issuing from the services and applications layer which consists in asking the system to launch a learning operation from a dataset stored in the Hadoop storage cluster, the Fig. 4 explain the explains the information exchange between the different components of the system in this case.

An explanation of the steps will be attempted using a sequence diagram illustrated in Fig. 5.

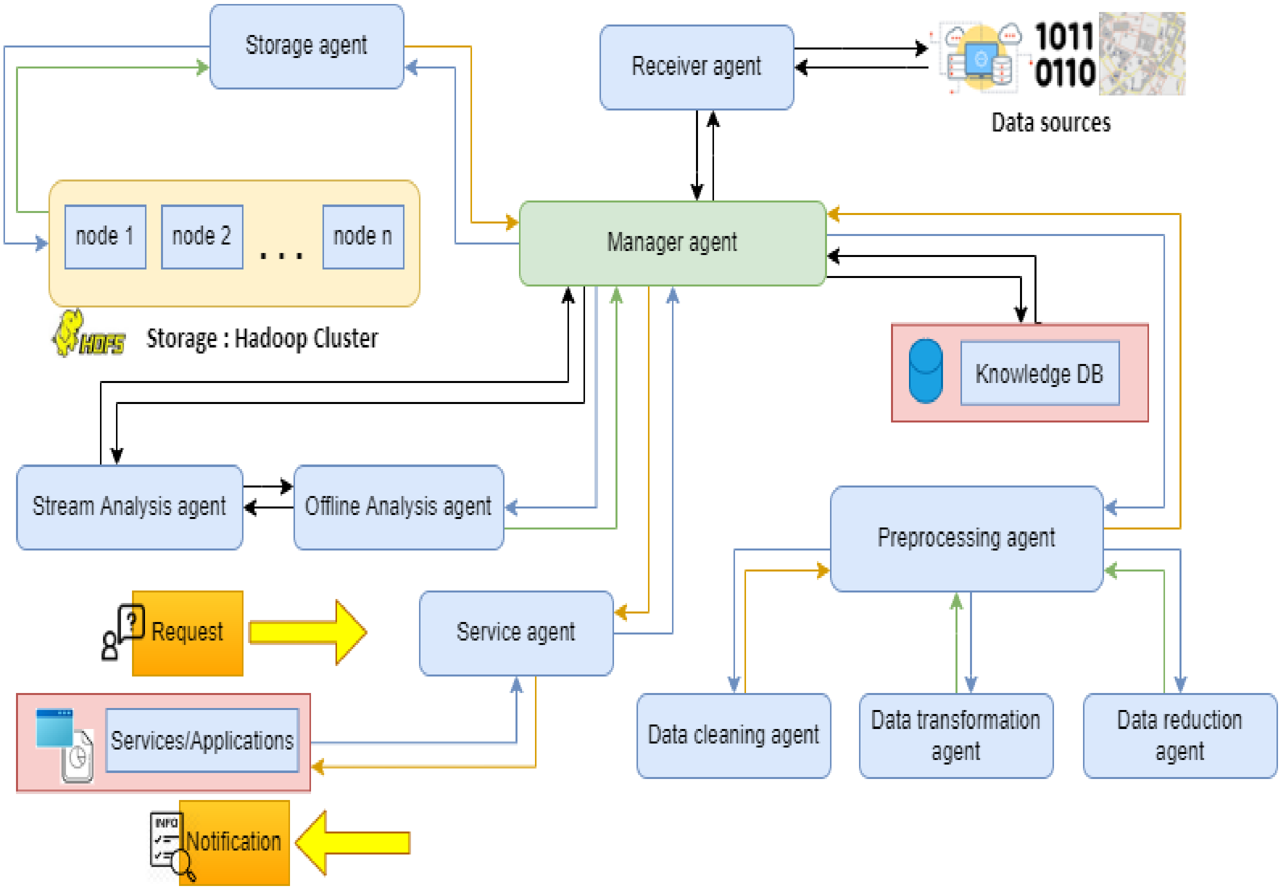


Fig. 4. The Flowchart of Training a ML Model using the Proposed MAS.

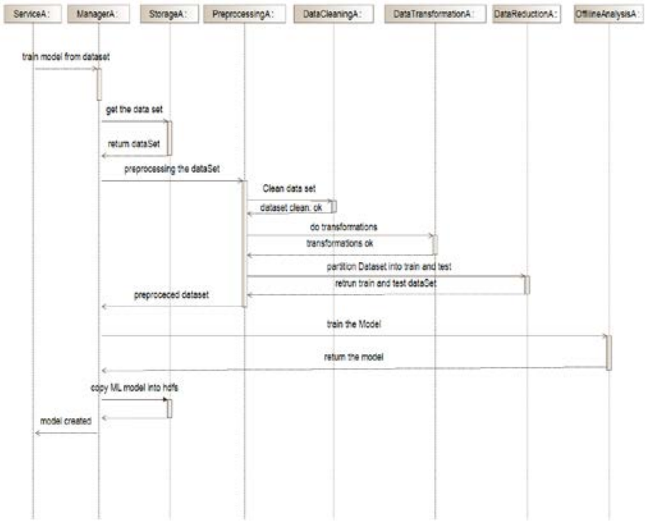


Fig. 5. Sequence Diagram of Training a ML Model using the Proposed MAS.

The service agent receives the request and sends it to the agent manager who is in charge of distributing the operations between agents, it sends the request to retrieve the Dataset to the storage agent, then it sends the Dataset for preprocessing operations by the Preprocessing Agent, which delegates tasks as necessary to the transformation, reduction and data cleaning agents.

Then the partitioned datasets are ready for the phases of training and testing by the Offline Analysis agent, once the model is valid the Manager agent sends the request to the storage agent to copy it on the storage cluster and notifies the service agent of the creation of the requested model.

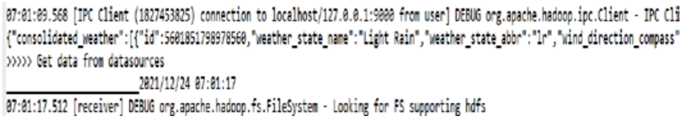
This generated model will be used to serve prediction requests that the system can receive and respond to in real time.

B. Case 2: Collecting Data from a Data Source and Storing it in HDFS

The proposed system can interact with several data sources such as databases, sensors, connected objects-IOT and external applications.

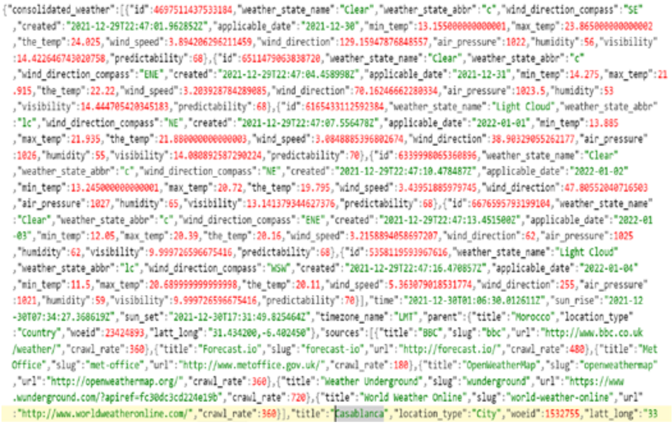
In this case the system is used to receive data from an external source which is an API[36] that provides weather data, the system opted to collect and store its data for the city of Casablanca in the HDFS file system.

In this example the agent will send request to the API in every specified lapse of time to get and store the last data of the weather about the city of Casablanca, the Fig. 6 represents a snapshot of a request of an example of a request sent by the system to get weather data.



```
07:01:09.568 [IPC Client (1827453825) connection to localhost/127.0.0.1:9000 from user] DEBUG org.apache.hadoop.ipc.Client - IPC Cli
{"consolidated_weather":[{"id":5601851798978560,"weather_state_name":"Light Rain","weather_state_abbr":"lr","wind_direction_compass"
>>>>> Get data from datasources
____________________2021/12/24 07:01:17
07:01:17.512 [receiver] DEBUG org.apache.hadoop.fs.FileSystem - Looking for FS supporting hdfs
```

Fig. 6. Agent's request to the API.



```
{"consolidated_weather":[{"id":4697511437533184,"weather_state_name":"Clear","weather_state_abbr":"c","wind_direction_compass":"SE"
 ,"created":"2021-12-29T22:47:01.962852Z","applicable_date":"2021-12-30","min_temp":13.155000000000001,"max_temp":23.865000000000002
 ,"the_temp":24.025,"wind_speed":3.894206296211459,"wind_direction":129.15947876848557,"air_pressure":1022,"humidity":56,"visibility"
 :14.422646743020758,"predictability":68},{"id":6511479063838720,"weather_state_name":"Clear","weather_state_abbr":"c"
 ,"wind_direction_compass":"ENE","created":"2021-12-29T22:47:04.458998Z","applicable_date":"2021-12-31","min_temp":14.275,"max_temp":21
 .915,"the_temp":22.22,"wind_speed":3.203928784289085,"wind_direction":70.16246662280334,"air_pressure":1023.5,"humidity":53
 ,"visibility":14.444705420345183,"predictability":68},{"id":6165433112592384,"weather_state_name":"Light Cloud","weather_state_abbr"
 :"lc","wind_direction_compass":"NE","created":"2021-12-29T22:47:07.556478Z","applicable_date":"2022-01-01","min_temp":13.885
 ,"max_temp":21.935,"the_temp":21.880000000000003,"wind_speed":3.0848885396802674,"wind_direction":38.90329055262177,"air_pressure"
 :1026,"humidity":55,"visibility":14.080892587290224,"predictability":70},{"id":6339998065360896,"weather_state_name":"Clear"
 ,"weather_state_abbr":"c","wind_direction_compass":"NE","created":"2021-12-29T22:47:10.478487Z","applicable_date":"2022-01-02"
 ,"min_temp":13.245000000000001,"max_temp":20.72,"the_temp":19.795,"wind_speed":3.43951885979745,"wind_direction":47.80552040716503
 ,"air_pressure":1027,"humidity":65,"visibility":13.141379344627376,"predictability":68},{"id":6676595793199104,"weather_state_name"
 :"Clear","weather_state_abbr":"c","wind_direction_compass":"ENE","created":"2021-12-29T22:47:13.451500Z","applicable_date":"2022-01
 -03","min_temp":12.05,"max_temp":20.39,"the_temp":20.16,"wind_speed":3.2158894058697207,"wind_direction":62,"air_pressure":1025
 ,"humidity":62,"visibility":9.999726596675416,"predictability":68},{"id":5358119593967616,"weather_state_name":"Light Cloud"
 ,"weather_state_abbr":"lc","wind_direction_compass":"WSW","created":"2021-12-29T22:47:16.470857Z","applicable_date":"2022-01-04"
 ,"min_temp":11.5,"max_temp":20.689999999999998,"the_temp":20.11,"wind_speed":5.363079018531774,"wind_direction":255,"air_pressure"
 :1021,"humidity":59,"visibility":9.999726596675416,"predictability":70}],"time":"2021-12-30T01:06:30.012611Z","sun_rise":"2021-12
 -30T07:34:27.368619Z","sun_set":"2021-12-30T17:31:49.825464Z","timezone_name":"LMT","parent":{"title":"Morocco","location_type"
 :"Country","woeid":23424893,"latt_long":"31.434200,-6.402450"},"sources":[{"title":"BBC","slug":"bbc","url":"http://www.bbc.co.uk
 /weather/","crawl_rate":360},{"title":"Forecast.io","slug":"forecast-io","url":"http://forecast.io/","crawl_rate":480},{"title":"Met
 Office","slug":"met-office","url":"http://www.metoffice.gov.uk/","crawl_rate":180},{"title":"OpenWeatherMap","slug":"openweathermap"
 ,"url":"http://openweathermap.org/","crawl_rate":360},{"title":"Weather Underground","slug":"wunderground","url":"https://www
 .wunderground.com/?apiref=fc30dc3cd224e19b","crawl_rate":720},{"title":"World Weather Online","slug":"world-weather-online","url"
 :"http://www.worldweatheronline.com/","crawl_rate":360}],"title":"Casablanca","location_type":"City","woeid":1532755,"latt_long":"33
```

Fig. 7. The Response of the API formatted as Json Object.

The Receiver agent can listen to changes in a data source or collect values from a data source according to the desired behavior e.g. in this case the Receiver agent adopts a TickBihaviour for each time lapse it retrieves the data and sends the request to the Manager agent to initiate the request for preprocessing to the Preprocessing agent, once the data is cleaned up and ready to be stored the Manager agent asks the Storage agent to copy it to the storage cluster. Fig. 7 shows an example of stored data: Name of the location, time of the location, name of the time zone in which the location is located, internal identifier of the forecast, date to which the forecast or observation refers, compass point of the wind direction, one or two letter abbreviation of the weather condition.

Fig. 8 tries to illustrate the functioning of the system using a diagram sequence.

C. Case 3: The use of a Machine Learning Model for Real-Time Prediction

In this case of study, the behavior of the system will be analyzed in the case where it must process prediction requests in real time.

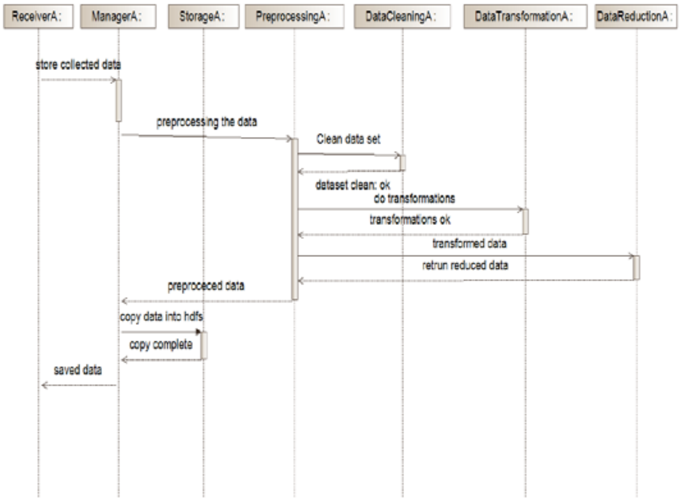


Fig. 8. Sequence Diagram of Collecting Data from a Data Source and Storing it in HDFS.

The first step is to build the model machine learning that will be used for future predictions, and to store it in the cluster, in this step a dataset will be used which contains data about households and individuals in Casablanca from the last general census of population and housing from HCP[37] this data set is made public in 2019, we conducted based on the analyze of this data a previous work to help in enhancing urban planning in the region of Casablanca by predicting the type of habitat and estimate the necessary equipment [23].

1) Training and test machine learning model: In the beginning of this learning phase the dataset used is already prepared cleaned and preprocessed, the start will be splitting the main dataset into two datasets usually in the literature the split is made into two part a larger one with 80% used to train the model and the part with 20% used to test the model, the Fig. 9 illustrate the steps of training the machine learning model used in this study.

The step of choosing the adequate algorithm is data-centric step based on the type of the data and the nature of the problem, in this study this dataset will be used to predict the type of habitat based on the characteristics of the studied population, it's a classification problem, in the next part a test of the main machine learning algorithms for supervised learning on the dataset will be performed.

We will present in this part the metrics and the obtained results for each used machine learning algorithm on the same dataset:

To evaluate the relative performance of different classification models, it is necessary to specify the appropriate quantitative criteria for the evaluation. The performance metrics are specified for the classifier:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Nagative} \quad (2)$$

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

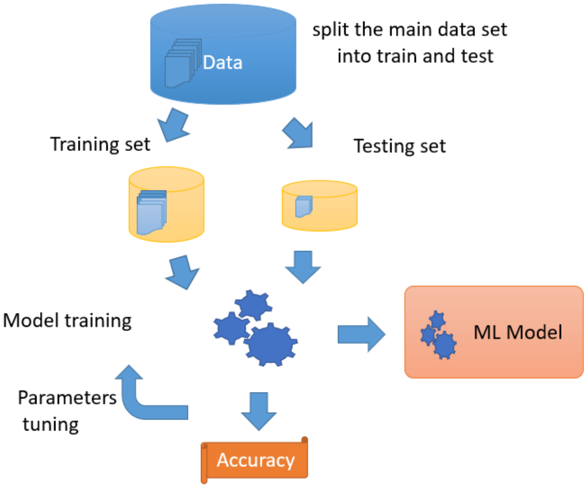


Fig. 9. Steps of Training ML Model using Classification Algorithms.

This section highlights the evaluation of the proposed model on the dataset for each classification algorithm.

a) Naïve bayes: A Naive Bayes classifier is a probabilistic machine learning model used in classification cases to discriminate different objects and labels based on a set of features, this classifier is based on the bayes theorem [38], with the naïve hypothesis of a conditional independence of each combination of features given the value of the class variable, the Table I presents the evaluation metric for the naïve bayes algorithm:

$$P(y|x_1, \dots, x_n) = \frac{P(y)P(x_1,\dots,x_n|y)}{P(y|x_1,\dots,x_n)} \quad (4)$$

TABLE I. CLASSIFICATION REPORT FOR THE ML ALGORITHM: NAÏVE BAYES

METRIC	SCORE (%)
precision	64
recall	61
f1-score	61

b) Multilayer perceptron: The MLP [39] multilayer perceptron is a type of artificial neural network organized as a set of multilayers, the process of information flowing from the input layer to the output layer only. It is therefore a feedforward network, each layer consists of a variable number of neurons, the neurons of the last layer being the outputs of the global system, The MLP are widely used in classification, prediction and recognition use cases, the Table II presents the metrics for the use of the multilayer perceptron classifier:

TABLE II. CLASSIFICATION REPORT FOR THE ML ALGORITHM MULTI LAYER PERCEPTRON

METRIC	SCORE (%)
precision	73
recall	73
f1-score	71

c) Random Forest: Random forest is a classifier that can be defined as a meta estimator which corresponds to a number of decision tree classifiers, this algorithm performs a splitting of the training data into a specific number of data subsets used to perform training on multiple decision trees, the subsets used for training the model are slightly different, and then a voting method is used to select the best model [40], the Table III shows the metrics for the use of the Random forest algorithm:

TABLE III. CLASSIFICATION REPORT FOR THE ML ALGORITHM RANDOM FOREST

METRIC	SCORE (%)
precision	85
recall	85
f1-score	84

d) Summary: The Table IV summarizes the evaluation metrics and give a comparison for the used classification algorithms.

TABLE IV. SUMMARY OF THE METRICS USING SUPERVISED LEARNING

METRIC	Precision	Recall	F1-Score	Accuracy
Naïve Bayes	0.64	0.61	0.61	0.61
MLP	0.73	0.73	0.71	0.73
Random Forest	0.85	0.85	0.84	0.85

To select the model to be deployed and used for future prediction requests some metrics are used like precision, recall, f1-score, and the accuracy.

Based on the results obtained, the most accurate model can be choose, in this case the model trained with Random Forest Algorithm, the next step will be to the step of serializing this model into the storage of the proposed system.

Serialization of the trained ML Model: In order to use the trained ML models in the different applications, the step of serialization still necessary, especially when the trained model is a result of huge dataset and machine learning algorithm that consumed a lot of resources to get to the validation phase of the model.

The reuse of the stored model still a good strategy that allows the system to use immediately the model, as explained in the Fig. 10.

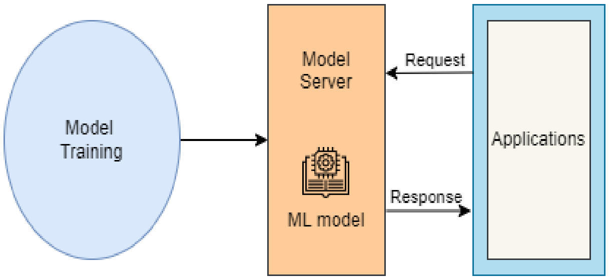


Fig. 10. Serialization and use of the Machine Learning Model.

The use of the trained model by the MAS system:

The proposed system can handle real-time prediction requests using the principle of collaboration between different agents, as shown in the Fig. 11 first the system get the request of prediction from the application and service layer this request is received with the service agent which will transmit it to the agent manager, the latter will call the storage agent to retrieve the serialized model in order to use it for this request and send the request for preprocessing to format it and to get the parameters or the datasets in this request, once done the operation of prediction by the stream analysis agent, this result will be sent to the service agent to bring up the response to the demanding application.

The system can handle the communication with different applications by defining an interface for communication with the proposed system,

And defining multiple functionalities to download, upload, store or analyze data by defining several functions.

For example, the receiver agent can receive requests of type:

$\text{Function}_i(\text{Arg}_1, \text{Arg}_2, \ldots, \text{Arg}_n) \Rightarrow D_i$

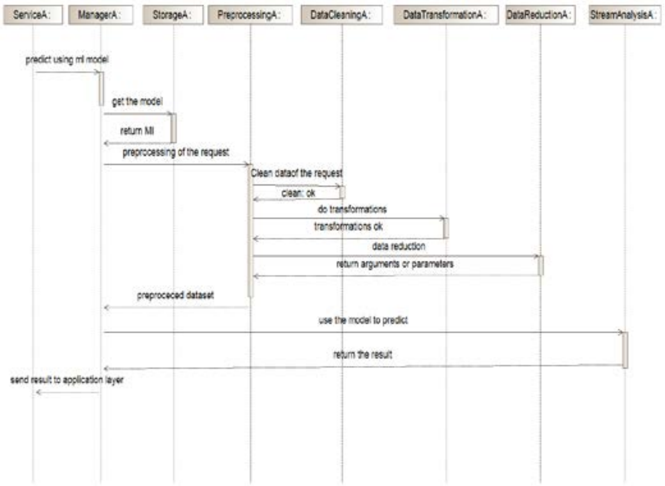
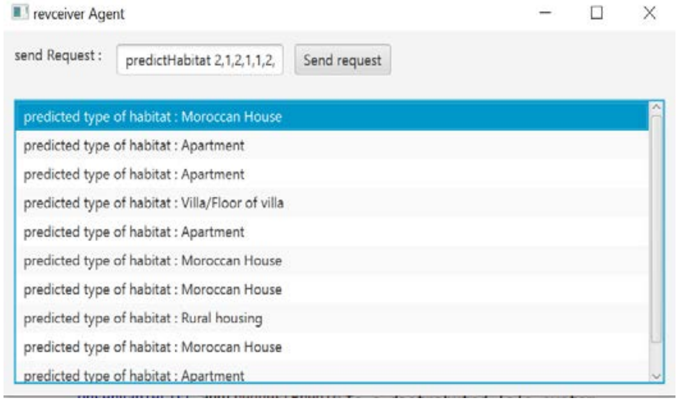


Fig. 11. Sequence Diagram of using a Machine Learning Model for Real-time Prediction.

These functions and services could be used in other applications like the one in the Fig. 12, which will interface with the proposed system and use these features.



predictHabitat 2,1,2,1,1,2,

predicted type of habitat : Moroccan House
predicted type of habitat : Apartment
predicted type of habitat : Apartment
predicted type of habitat : Villa/Floor of villa
predicted type of habitat : Apartment
predicted type of habitat : Moroccan House
predicted type of habitat : Moroccan House
predicted type of habitat : Rural housing
predicted type of habitat : Moroccan House
predicted type of habitat : Apartment

Fig. 12. Example of an Application Consuming the Prediction Service of the Proposed System.

VI. CONCLUSION AND PERSPECTIVES

In this paper, a multi-agent system to automate big data analytics in real time based on the smart data approach was proposed, a model with a multilayer architecture was presented, and the adaptive behavior of this system was illustrated through some case studies, especially in the case of training and the using machine learning models.

The use of intelligent processing, automating big data analytics through the use of the multi-agent system and machine learning techniques can help to solve complex problems such in the field of Urban planning and management in smart cities, by providing actionable data and efficient decisions.

To summarize, the paper tries to highlight the challenges following the processing of big data specifically issued from smart cities, these data are in different formats (structured, semi-structured, unstructured). In addition, they are generated at different rates by batch or in real time in streams. These voluminous amounts of data present many challenges including the management of these data, the latency of processing time and mainly the effective exploitation of these data to extract useful information.

To overcome these problematic, the paper proposes a model based on a multi-agent system that adopts the smart data approach which focuses on the value aspect of data and the application of intelligent processing in order to extract useful information from these voluminous masses of big data and to drive an efficient decision making, the proposed system automates its data processing processes, and thanks to the autonomous and cognitive capabilities of the agents they can manage and make decisions in real time, And to illustrate the adaptive behavior of the system through the autonomous behavior of the agents, we proposed 3 scenarios:

- The system builds a machine learning model of the classification from a large dataset that contains the demographic data of the city of Casablanca from the last census in Morocco in 2019, to predict the type of housing adequate to a profile of household or citizen for this several tasks were performed by the agents including the preprocessing phase and the choice of appropriate classification algorithm according to the study of metrics and then the serialization of the model, in order to use it for future requests for prediction in real time.
- The system listens to the various data sources and stores the data in the Hadoop storage cluster in real time.
- The system responds to real-time prediction demands by using machine learning models already built and stored in the system to improve the real-time decision-making process.

Such a proposition can find applications in several areas including the management and planning of smart cities known with the generation of huge amounts of data, this proposition can contribute in the effective use of these data and the extraction of valuable information that can guide effective decision-making to improve for example the management of traffic flows and road traffic, the anticipation of needs in terms of equipment in smart cities, reduce energy consumption, the monitoring of KPIs, the reporting and monitoring of smart cities.

In perspective, we will work on enhancing the system with other functionalities, focusing on the improvement of agent behaviors and the machine learning models. To test this approach, solve several problems in urban planning and management in smart cities.

The proposed approach in this paper represents a real opportunity to enhance the process of big data analytics by the exploit of the cognitive functionalities of the agents and to empower their capabilities using a knowledge base which will facilitates the process of decision- making made by agents to take actions in a autonomous way, such adaptive behavior can solve serious problems in smart cities like energy consumption and traffic congestion.

References

[1] Lee, “Big data: Dimensions, evolution, impacts, and challenges,” Business Horizons, vol. 60, no. 3, pp. 293–303, May 2017, doi: 10.1016/j.bushor.2017.01.004.

[2] F. Iafrate, “A Journey from Big Data to Smart Data,” in Digital Enterprise Design & Management, 2014, pp. 25–33.

[3] J. Luengo, D. García-Gil, S. Ramírez-Gallego, S. García, and F. Herrera, “Smart Data,” in Big Data Preprocessing, Cham: Springer International Publishing, 2020, pp. 45–51.

[4] F. Sassite, M. Addou, and F. Barramou, “A smart data approach for Spatial Big Data analytics,” in 2020 IEEE International conference of Moroccan Geomatics (Morgeo), May 2020, pp. 1–6, doi: 10.1109/Morgeo49228.2020.9121920.

[5] S. García, J. Luengo, and F. Herrera, “Tutorial on practical tips of the most influential data preprocessing algorithms in data mining,” Knowledge-Based Systems, vol. 98, pp. 1–29, Apr. 2016, doi: 10.1016/j.knosys.2015.12.006.

[6] A. Lenk, “Smart Data: Von Technologien zu Standards,” 2015.

[7] D. Agrawal, S. Das, and A. El Abbadi, “Big data and cloud computing: current state and future opportunities,” in Proceedings of the 14th International Conference on Extending Database Technology - EDBT/ICDT ’11, Uppsala, Sweden, 2011, p. 530, doi: 10.1145/1951365.1951432.

[8] B. R. Prasad and S. Agarwal, “Comparative Study of Big Data Computing and Storage Tools: A Review,” International Journal of Database Theory and Application, vol. 9, no. 1, pp. 45–66, Jan. 2016, doi: 10.14257/ijdta.2016.9.1.05.

[9] Y. K. Gupta and C. Jha, “A review on the study of big data with comparison of various storage and computing tools and their relative capabilities,” International Journal of Invocation in engineering & technology (IJIET), vol. 7, no. 1, pp. 470–477, 2016.

[10] D. García-Gil, J. Luengo, S. García, and F. Herrera, “Enabling Smart Data: Noise filtering in Big Data classification,” Information Sciences, vol. 479, pp. 135–152, Apr. 2019, doi: 10.1016/j.ins.2018.12.002.

[11] F. SASSITE, M. ADDOU, and F. BARRAMOU, “A Smart Data Approach for Automatic Data Analysis,” in ESAI’19: 1St international conference on embedded systems and artificial intelligence, 2019.

[12] A. Dorri, S. S. Kanhere, and R. Jurdak, “Multi-Agent Systems: A Survey,” IEEE Access, vol. 6, pp. 28573–28593, 2018, doi: 10.1109/ACCESS.2018.2831228.

[13] N. Benmoussa, M. Fakhouri Amr, S. Ahriz, K. Mansouri, and E. Illoussamen, “Outlining a Model of an Intelligent Decision Support System Based on Multi Agents,” Engineering, Technology & Applied Science Research, vol. 8, no. 3, pp. 2937–2942, Jun. 2018, doi: 10.48084/etasr.1936.

[14] M. Naserian, A. Ramazani, A. Khaki-Sedigh, and A. Moarefianpour, “Fast terminal sliding mode control for a nonlinear multi-agent robot system with disturbance,” Systems Science & Control Engineering, vol. 8, no. 1, pp. 328–338, Jan. 2020, doi: 10.1080/21642583.2020.1764408.

[15] F. L. Bellifemine, G. Caire, and D. Greenwood, Developing multi-agent systems with JADE. 2010.

[16] P. P. Shinde and S. Shah, “A Review of Machine Learning and Deep Learning Applications,” in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Aug. 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697857.

[17] J. A. Nichols, H. W. Herbert Chan, and M. A. B. Baker, “Machine learning: applications of artificial intelligence to imaging and diagnosis,” Biophysical Reviews, vol. 11, no. 1, pp. 111–118, Feb. 2019, doi: 10.1007/s12551-018-0449-9.

[18] N. C. Eli-Chukwu, “Applications of Artificial Intelligence in Agriculture: A Review,” Engineering, Technology & Applied Science Research, vol. 9, no. 4, pp. 4377–4383, Aug. 2019, doi: 10.48084/etasr.2756.

[19] M. Anwer, S. M. Khan, M. U. Farooq, and W. Waseemullah, “Attack Detection in IoT using Machine Learning,” Engineering, Technology & Applied Science Research, vol. 11, no. 3, pp. 7273–7278, Jun. 2021, doi: 10.48084/etasr.4202.

[20] D. Xia, P. Chen, B. Wang, J. Zhang, and C. Xie, “Insect Detection and Classification Based on an Improved Convolutional Neural Network,” Sensors, vol. 18, no. 12, p. 4169, Nov. 2018, doi: 10.3390/s18124169.

[21] P. Ghadimi, C. Wang, M. K. Lim, and C. Heavey, “Intelligent sustainable supplier selection using multi-agent technology: Theory and application for Industry 4.0 supply chains,” Computers & Industrial Engineering, vol. 127, pp. 588–600, Jan. 2019, doi: 10.1016/j.cie.2018.10.050.

[22] E. Erturk and K. Jyoti, “Perspectives on a Big Data Application: What Database Engineers and IT Students Need to Know,” Engineering, Technology & Applied Science Research, vol. 5, no. 5, pp. 850–853, Oct. 2015, doi: 10.48084/etasr.592.

[23] F. Sassite, M. Addou, and F. Barramou, “Towards a Multi-agents Model for Automatic Big Data Processing to Support Urban Planning,” in Geospatial Intelligence, F. Barramou, E. H. El Brirchi, K. Mansouri, and Y. Dehbi, Eds. Cham: Springer International Publishing, 2022, pp. 3–17.

[24] G. Lombardo, P. Fornacciari, M. Mordonini, M. Tomaiuolo, and A. Poggi, “A Multi-Agent Architecture for Data Analysis,” Future Internet, vol. 11, no. 2, p. 49, Feb. 2019, doi: 10.3390/fi11020049.

[25] K. Dounya, K. Okba, S. Hamza, and B. Omar, “Design and Implementation of a New Approach using Multi-Agent System for Security in Big Data,” International Journal of Software Engineering and Its Applications, vol. 11, no. 9, pp. 1–14, Sep. 2017, doi: 10.14257/ijseia.2017.11.9.01.

[26] C. Yao, S. Wu, Z. Liu, and P. Li, “A deep learning model for predicting chemical composition of gallstones with big data in medical Internet of Things,” Future Generation Computer Systems, vol. 94, pp. 140–147, May 2019, doi: 10.1016/j.future.2018.11.011.

[27] A. Shashwat and D. Kumar, “A service identification model for service oriented architecture,” in 2017 3rd International Conference on Computational Intelligence Communication Technology (CICT), Feb. 2017, pp. 1–5, doi: 10.1109/CIACT.2017.7977299.

[28] T. H. Akila, S. Siriweera, I. Paik, and B. T. G. S. Kumara, “Onotology-based service discovery for intelligent Big Data analytics,” in 2015 IEEE 7th International Conference on Awareness Science and Technology (iCAST), Sep. 2015, pp. 66–71, doi: 10.1109/ICAwST.2015.7314022.

[29] B. T. G. S. Kumara, I. Paik, J. Zhang, T. H. A. S. Siriweera, and K. R. C. Koswatte, “Ontology-Based Workflow Generation for Intelligent Big Data Analytics,” in 2015 IEEE International Conference on Web Services, New York, NY, USA, Jun. 2015, pp. 495–502, doi: 10.1109/ICWS.2015.72.

[30] I. Paik, W. Chen, and M. N. Huhns, “A Scalable Architecture for Automatic Service Composition,” IEEE Transactions on Services Computing, vol. 7, no. 1, pp. 82–95, Jan. 2014, doi: 10.1109/TSC.2012.33.

[31] A. Gandomi and M. Haider, “Beyond the hype: Big data concepts, methods, and analytics,” International Journal of Information Management, vol. 35, no. 2, pp. 137–144, Apr. 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.

[32] J. Luengo, D. García-Gil, S. Ramírez-Gallego, S. García, and F. Herrera, “Big Data: Technologies and Tools,” in Big Data Preprocessing, Cham: Springer International Publishing, 2020, pp. 15–43.

[33] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, “Big data preprocessing: methods and prospects,” Big Data Analytics, vol. 1, no. 1, p. 9, Dec. 2016, doi: 10.1186/s41044-016-0014-0.

[34] D. Pyle, Data preparation for data mining. morgan kaufmann, 1999.

[35] M. K. Eddy, A. Ahmad, and A. Y. C. Tang, “Agents of Things (AOT): Utilizing JADE Agent Technology as Communication Middleware for Vehicle Monitoring System,” International Journal of Future Generation Communication and Networking, vol. 11, no. 1, pp. 47–55, Jan. 2018, doi: 10.14257/ijfgcn.2018.11.1.05.

[36] “API - MetaWeather.” https://www.metaweather.com/api/location/1532755/ (accessed Dec. 30, 2021).

[37] “RGPH 2014 | Téléchargements | Site institutionnel du Haut-Commissariat au Plan du Royaume du Maroc,” Nov. 28, 2020. https://www.hcp.ma/downloads/RGPH-2014_t17441.html (accessed Nov. 28, 2020).

[38] H. Zhang, “The Optimality of Naïve Bayes,” in In FLAIRS2004 conference, 2004.D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv:1412.6980 [cs], Jan. 2017, Accessed: Mar. 21, 2022. [Online]. Available: http://arxiv.org/abs/1412.6980.

[39] A. Sarica, A. Cerasa, and A. Quattrone, “Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer’s Disease: A Systematic Review,” Frontiers in Aging Neuroscience, vol. 9, 2017, doi: 10.3389/fnagi.2017.00329.

Brain Tumor Detection using MRI Images and Convolutional Neural Network

Driss Lamrani[1], Bouchaib Cherradi[2], Oussama El Gannour[3], Mohammed Amine Bouqentar[4], Lhoussain Bahatti[5]
EEIS Laboratory, ENSET of Mohammedia, Hassan II University of Casablanca, Mohammedia, Morocco[1,2,3,4,5]
STIE Team, CRMEF Casablanca-Settat, provincial section of El Jadida, El Jadida, Morocco[2]

***Abstract*—A brain tumor is the cause of abnormal growth of cells in the brain. Magnetic resonance imaging (MRI) is the most practical method for detecting brain tumors. Through these MRIs, doctors analyze and identify abnormal tissue growth and can confirm whether the brain is affected by a tumor or not. Today, with the emergence of artificial intelligence techniques, the detection of brain tumors is done by applying the techniques and algorithms of machine learning and deep learning. The advantages of the application of these algorithms are the quick prediction of brain tumors, fewer errors, and greater precision, which help in decision-making and in choosing the most appropriate treatment for patients. In the proposed work, a convolution neural network (CNN) is applied with the aim of detecting the presence of a brain tumor and its performance is analyzed. The main purpose of this article is to adopt the approach of convolutional neural networks as a machine learning technique to perform brain tumor detection and classification. Based on training and testing results, the pre-trained architecture model reaches 96% in precision and classification accuracy rates. For the given dataset, CNN proves to be the better technique for predicting the presence of brain tumors.**

Keywords—Brain tumor; machine learning; convolutional neural network; MRI images

I. Introduction

A brain tumor is one of the major health problems related to human brain abnormalities [1]–[3]. It is a collection or a combination of unnatural tissue in the brain [4]. It usually occurs due to the fast growth of abnormal or damaged cells and created lamps of tissue in the brain. To find out whether the brain contains a tumor or not, MRI images are the first step in medical diagnosis [5], followed by manual analysis by an expert who looks for lesions in the brain. MRI is the most common and widely used technique for brain tumor detection and brain imaging as it proves better results especially when it It's about detecting differences between body tissues [6]. Compared to the computerized tomography (CT) scan and other techniques, MRI images are safer and can produce higher contrast images of brains [4]. Likewise, due to the high resolution of the images provided by MRI scans, a lot of information about the brain structure and the brain tissues are easily provided, this has considerable advantages in the field of image analysis [7], [8]. There are four standard sequences used in MRI scans modalities: The T1-weighted MRI (T1), T2-weighted MRI (T2), T1-weighted contrast-enhanced MRI (T1-CE), and FLAIR [9], [10]. In general, the T1W contrast-enhanced is the most commonly used modality, because it allows an easy annotation of the healthy tissues [11], [12].

Deep learning is a type of artificial intelligence derived from machine learning. Unlike programming where the content to execute the predetermined rules, here the machine can learn by itself, relying on a network of artificial neurons inspired by the human brain. Due to its higher performance in several fields as screening medical face mask [13], image description and a lot of challenges, the exploitation of DL technique in the medical image for classification, detection, and segmentation is highly encouraged [14]. In fact, various human diseases could be detected using such techniques, including COVID-19 [15]–[19], Parkinson's disease [20], breast cancer [21], diabetes diseases [22], medical image segmentation [23], and heart disease prediction [24]–[26]. A vast range of different scientific topics has developed because of advances in AI [27]–[35].

Convolutional neural networks (CNNs) are deep neural networks that have the capability to classify and segment images [36], [37]. CNN architectures for classification and segmentation include a variety of different layers with specific purposes, such as a convolutional layer, pooling layer, fully connected layers, dropout layers, etc. Recent studies on deep learning with convolutional neural networks have achieved excellent performance that is almost at the same level of performance of practicing radiologists [38], [39].

In recent years, CNN has been used in medical image segmentation. It has achieved great success in the field and auxiliary diagnosis. Compared to the traditional methods, network architecture gives to CNN algorithms the ability to learn complicated features from images [40], also CNNs are the top ranked and most popular algorithms used in computer vision, and actively contribute in this specific area of imaging due to their capability of detecting significant features, more particularly when it comes to medical imaging in which to processing difficult images with several details is needed, indeed CNN-based approaches are placed in the well-used leader board of the many image understanding challenges, especially Brain Tumor segmentation (BRATS), the biomedical challenge of Image Computing and Computer Assisted Intervention (MICCAI), ISBI (International Symposium on Biomedical Imaging), and IPMI (Information Processing in Medical Imaging) [41]. In this context, an application of a machine learning algorithm based on convolutional neural networks for the detection and classification of brain tumors is proposed.

The remainder of this paper is organized as follows: Section II presents a brief review of related works. Section III explains our proposed methodology and the machine learning algorithms implementation. Some performance evaluation techniques are presented in Section IV. Results and discussion are given in Section V. Section VI presents a conclusion with a discussion regarding the future works.

II. Related Work

In the research published in recent years relating to the segmentation of brain tumors, there is two branches of machine learning methods: Supervised learning and unsupervised learning .The difference between these models is simple, in supervised learning human intervention is needed at least to label the data appropriately, the machine learning model learns through iterating the operations of prediction of the labeled data communicated as input and improves the results by adjusting the response each time [42].

In unsupervised method-based segmentation, the human touch always needed for validating output variables [43], among its techniques there is the support vector machine (SVM) method and the fuzzy clustering approach are used. These methods have achieved the best performance when it comes to predicting or detecting a tumor, on the other hand the performance obtained is a little weak when the border between the normal tissue and the tumor area is blurred. Also, the extraction process usually takes time because the extraction algorithms have to extract a lot of information and details related to these edges and a lot of essential features.

Segmentation is the process of separating the image into distinct regions and is one of the most vital and demanding aspects of computer-aided clinical diagnostic tools. Although brain tumor segmentation is primarily done manually, it is very time consuming and the segmentation is subject to variations both between observers [44]. Therefore, an automatic and robust brain tumor segmentation will have a significant impact on brain tumor diagnosis and treatment.

Convolutional neural networks (CNN) differ from other machine learning techniques in the fact that the segmentation is done automatically and does not require the intervention of an expert as well as the feature extraction which is done with precision, on the other hand there are several parameters to be learned by a neural network (CNN) which results in expensive computation time and requires process graphics units (GPU) to train the model [45]. The role of a CNN network consists of two main tasks: feature extraction and data classification, the convolution and pooling layer are responsible for feature extraction, fully connected layers help to achieve the classification task [46]. In [47], the authors presented a convolutional neural network for brain tumor detection, 2 CNNs models were made as a comparison to find the best model for classification. The first model includes one convolution layer, the second includes two layers. This study showed that increasing convolutional layers improves the model performances and resulted in accuracy of 93% and a loss value of 0,23.

Among the techniques for using artificial neural networks, there is also transfer learning, the principle is simple, instead of designing a new CNN model and training it from scratch, existing architectures which are well trained on large dataset and which have demonstrated their performance are used, this makes it possible to use each transfer learning model and adapt it to the desired stain according to the nature of the task and the characteristic to be detected or classified [48].

Due to their ability to self-learn without the intervention of an expert, CNN models based on Transfer learning techniques have achieved excellent performance, the use of the weight sharing technique provides an adequate network and allows to automatically detect the tumor through the MRI images [49].

A research published in [50], Aimed boosting accuracy by the use of transfer learning strategy they implemented three transfer learning approaches using pre-trained CNN models, namely: VGG19 ,Inception V3 and MobileNet V2, They obtained respectively an accuracy rate of 88,22%, 91% and 92%. The authors concluded that the MobilnetV2 is the most efficient Model compared to the other models.

Using the ANN approach, Authors in [51],worked on two classes named benign tumors and malignant tumors, they started by preprocessing the images with the filters, then applied the average color moment technique on the images to extract the characteristics. After the transmission of these characteristic maps to the ANN, the classification was made with an accuracy rate of 91.8%.

A study published in [52], the model uses the histogram statistical equalization technique which consists in applying a transformation to each pixel of the image by calculating several statistical characteristics such as the average sum, the variance, the entropy, the dissimilarity, this model is therefore used for low-grade and high-grade class images of cervical glioma. The results obtained from the proposed method of accuracy, sensitivity and specificity reached 83.6% accuracy, 80.88% sensitivity and 86.84% specificity.

One of the techniques used in the detection of characteristics and classification of images consists in combining the concept of deep learning, CNNs, with other methods of preprocessing, these techniques include data augmentation, edge detection, genetic algorithm (GA), discrete wavelet transform (DWT) and principal component analysis (PCA). Authors in [53] have combined the two techniques of data augmentation and edge detection, data augmentation makes it possible to increase the amount of data artificially, other images are generated from the first images provided. The edge detection will allow finding the region of interest (ROI), the extraction of the characteristics is done thanks to a simple CNN model. They obtained 89% classification accuracy.

The results of these searches are classified in Table I, highlighting the classification model with a description of the chosen preprocessing technique, as well as the scores obtained based on the metrics used in each work.

TABLE I. SUMMURY OF SOME RELATED WORKS

Study	Dataset	Processing technique	Classifier	Result

[47]	Kaggle Dataset	Data Augmentation	CNN	Ac=93%
				Loss = 0,23

[50]	Brain Tumor Dataset	Transfer learning	VGG19	Ac=88%
				F1score=88.18%

			Inception V3	Ac=91%
				F1score=90.98%

			MobileNetV2	Ac=88%
				F1score=88.18%

[51]	Harvard Medical School Website Dataset	Noise Filtering Average color Moment	ANN	Ac=91.8%

[52]	LGG Flair MRI images	ROI feature extraction	Random forest	Ac=83.6%
				Sensitivity=80.88%
				Specificity=86.84%

III. MATERIALS AND METHODOLOGY

A. Global Overview on Proposed Detection Model

To complete the image classification task using machine leaning techniques such as conventional neural networks, these steps must be followed in sequential order: data extraction, data preprocessing, Feature selection, learning and classification Considering the common standard process of machine learning, our first step concerned the collection of data, the data is retrieved from Kaggle datasets of Brain MRI Images [54] The data folder concerns a set of 3000 MRI data images classified according to the presence or not of the tumor and labelled (Yes tumor and No tumor). In the development of our CNN architecture, we took into consideration several criteria to avoid limitations that are not suitable for our case study. One of the major limitations is the limited data, medical images are difficult to retrieve for privacy reasons and other several reasons, to have a robust model a large number of medical images is needed; there is currently some techniques which can limit this problem like Data augmentation. Several technical considerations are also considered. For example, the use of complicated transforms which can disturb the learning process is avoided, according to some research results, it was found that complicated transforms are not always better than simple ones because they can introduce some noise in the features and disturbs the learning process.

For the data pre-processing step, image partitioning techniques as well as the normalization of their size are involved. After this stage, the CNN is defined and implemented model as an algorithm for brain tumor detection. Then the input data is divided into training, validation, and test data. Having defined and compiled the model, evaluation metrics algorithm is defined in order to evaluate the model performance, then some predictions are made by executing the model on some MRI images. Fig. 1, describes the CNN architecture model implemented in our work, Layers definitions and roles are described in Subsection D.

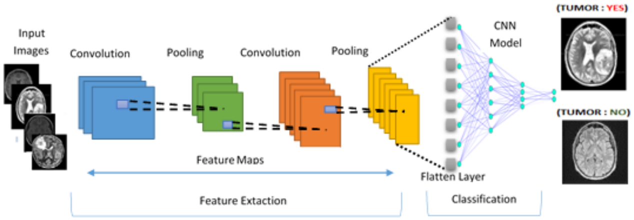


Fig. 1. Architecture of the Proposed Model.

B. Dataset Collection and Preprocessing

This section present some information about the database used, in fact a publicly accessible Kaggle database [54] is used . This database is composed of three folders, the first folder contains 1500 MRI scans presenting a brain tumor, the second folder contains 1500 MRI of healthy brains, in addition there is a folder containing some unlabeled MRI scans for testing purpose, the latter is not used, because another approach for the test data is planned. Thus, the final database obtained is built on the first 2 files and consists of 3000 images as input data distributed as follows: 1500 images with tumor and 1500 images without tumor. Fig. 2 illustrates the sample images of the dataset.

MRI images are not necessarily clear, sometimes visualization defects the quality of the image, these defects result from poor quality of image distortion and resolution, and could lead to a false analysis, and affect patient treatment options. Therefore, several preprocessing techniques can be introduced to make the images more robust and more usable by the neural network, the most common techniques concern the aspect ratio: uniform image size, dimensionality reduction, and data augmentation. The images have been resized to (224, 224, 3) = (image width, image height, number of channels) to facilitate the learning process.

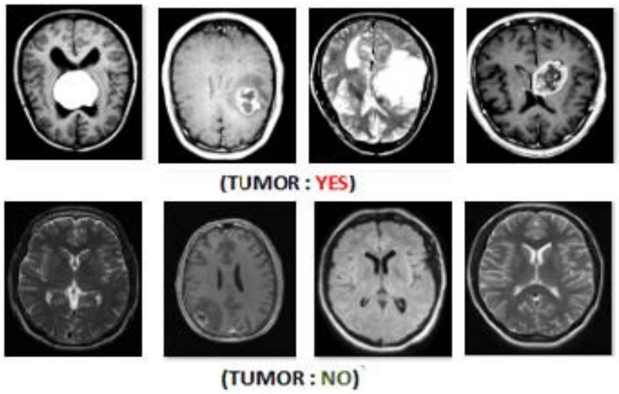


Fig. 2. Samples Images for Brain Tumor Dataset.

Input image will be processed into the first pre-process which is the process of wrapping and cropping, it consists of the removal of unwanted outer areas and some of the peripheral areas from a photographic or illustrated image, after that the database is divided into training, testing, and validation sets with an 80:10:10 ratio. The 3000 images dataset is split into training, testing, and validation sets of 2398, 300, and 300 images respectively, and then all the images are processed in the collection into an array. The last step is the coding when the tagged data are transformed into a numerical label so that they may be interpreted and analyzed. Fig. 3 describes the flowchart followed by the model from input images and preprocessing to the CNN model algorithm and the prediction of healthy and unhealthy brains.

C. Machine Learning Algorithm for Classification

Classification algorithms classify brain tumors into respective categories. It has an essential task in interpreting, extracting features, analyzing, and interpreting images in many applications. The CNN model must first extract features from each image before learning how to distinguish between the images provided to it. In this research, a CNN model with several layers is proposed; four convolution layers, three Maxpooling layers, one flatten layer, and six dense layers. In general, the core building blocks of convolutional neural networks are convolutional layers, activation functions, pooling, and fully connected layers. To improve the results in the output of the CNN network, the input data must go through several stages. The main objective is to correct the adjustments, allowing the CNN to recognize the inaccurate features in the images. The over-fitting correction is done through four techniques: data augmentation, dropout, batch normalization, and pooling. These process steps represent the hidden layers of the neural network and are used to perform the CNN model. Some definitions and roles of these layers are described below.

To build our proposed model, the first required Conv2D parameter is the number of filters that the convolutional layer will learn. In the proposed architecture, the input is an image of (224*224*3) size, 20 filters are implemented with a kernel size of (4*4). The same parameters are kept for the other two convolution layers with a reduction in kernel size to (2*2). After each convolution operation, the Maxpooling layers with successive calls to the dense layers and the RELU activation function are introduced. In the dense layers, the unit's values are respectively (1024, 512, 256, 128, 64). Finally, the Softmax function is used as the activation function in the output layer. There are 14 547 134 parameters in total. All these details are described in Fig. 4, which illustrates the architecture of our proposed model.

- Convolutional layer Is the main layer of a convolutional neural network, and composed of a filter for the input data, a feature map and a feature detector known as the CNN core whose role is the detection and the extracting of the features in the image, such as edges and colors [55].
- The Pooling layer Thanks to the spatial variance property, Maxpooling teaches the neural network that it is the same characteristic to be detected despite the differences that may exist between the images of the same object, namely the way in which the images are presented, dimensions, textures. This can only be done after having a ready features map.
- Flatten Layer represents the input layer for the artificial neural network, this phase consists of grouping an entity map in a column, and hence the name "Flattening", this allows to have a large data vector compatible with the neural network input.
- Max-Pooling Layer: Maxpooling is used to extract the most relevant features in the images, in the original entity map, the end is always the maximum value, and unnecessary details in each image have are removed to allow network of neurons to do the job efficiently.

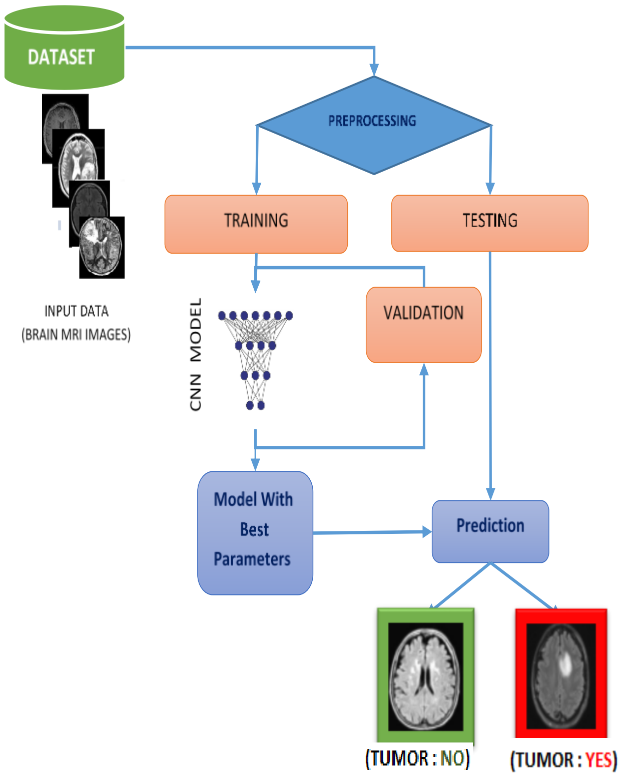


Fig. 3. The Flowchart to Implementation of the CNN Model.

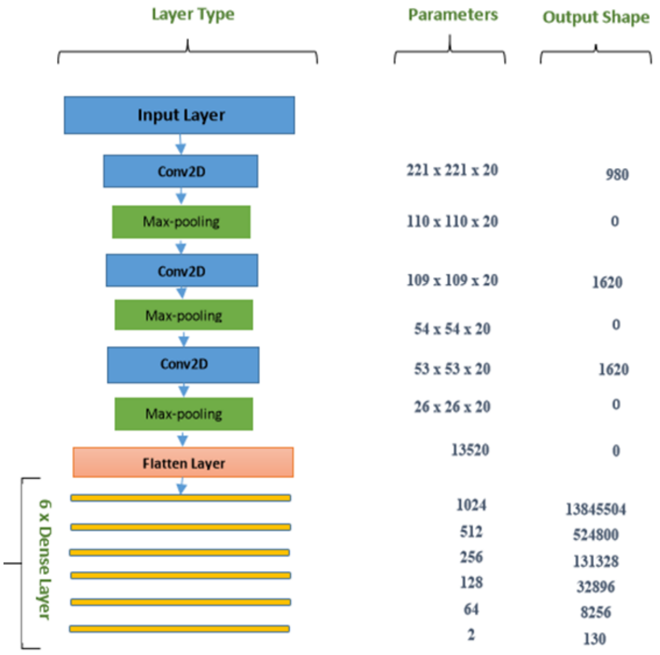


Fig. 4. The Proposed Architecture for CNN Model.

IV. Performance Evaluation Metrics

Several standard evaluation metrics are generally introduced to evaluate the performance of the system, among these metrics there is Accuracy, Precision, Recall, AUC, F1_score, confusion matrix and Receiver Operating Characteristic curve (ROC). The principle of metrics and their mathematical calculation formula is detailed below.

- Confusion Matrix

A confusion matrix is represented by a two-dimensional table which summarizes the results of the predictions of the classification carried out and allows to compare between the correct and false results of the prediction, which allows to see at what point a model can be confused in its predictions and to measure these performances. In a confounding matrix the results are classified into four main categories: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The four elements of the confusion matrix are shown in Table II.

TABLE II. Elements of the Confusion Matrix

Element	Description
TP	Images containing the tumor and correctly classified.
NP	Images not containing the tumor and correctly classified.
FP	CNN classifies images as containing tumors but that does not contain any tumor.
FN	CNN classifies images as not containing any tumor but are containing a tumor.

- Loss Function

Loss function is a function that evaluates the difference between the predictions made by the neural network and the actual values of the observations used during learning. The more the result of this function is minimized, the more the neural network performs. Its minimization, reducing to a minimum the difference between the predicted value and the actual value for a given observation, is done by adjusting the different weights of the neural network.

V. Results and Discussion

Several standard evaluation metrics are generally introduced to evaluate the performance of the system, among these metrics there is Accuracy, Precision, Recall, AUC, F1_score, confusion matrix and Receiver Operating Characteristic curve (ROC).

A. Training Results

The CNN model is trained using a notebook from the open Google Colab platform which allows to take full advantage of popular Python libraries like TensorFlow and Keras. Colab notebooks run this code on Google's cloud servers and put in our service the power of Tesla K80 GPU with 12 GB of GDDR5 VRAM, increasing performance and reducing training time considering the large number of parameters to train with the proposed CNN model .in addition this platform had an Intel Xeon Processor with two cores running at 2.20 GHz, and 13 GB of RAM. Fig. 5 and Fig. 6 show the evolution of the accuracy and loss curve during the two training and validation periods.

From the plot of Accuracy curves in Fig. 5 which represents the graph evolution during the training period and the validation period, it is clearly seen that the training accuracy is higher than the validation accuracy; the same for the loss curve in Fig. 6, the training curve is above the validation curve which leads to the conclusion that the proposed CNN model has no overfitting issue.

B. Testing Results

The performance of the proposed methodology was evaluated by the measures of precision, specificity, accuracy and, above all, the ROC curve for two classes (normal and abnormal brains) and compared to the performance of other classifiers, the metrics mathematical equations are detailed below.

$$Specificity = \frac{TN}{TN+FP} \tag{1}$$

$$Sensivity = \frac{TP}{FN+TP} \tag{2}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$F1 - Score = 2 * \frac{Precision * Sensivity}{Precision+Sensivity} \tag{5}$$

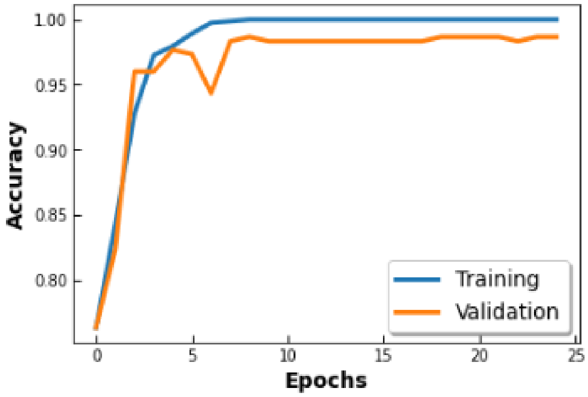


Fig. 5. Accuracy Curve of the Proposed Model.

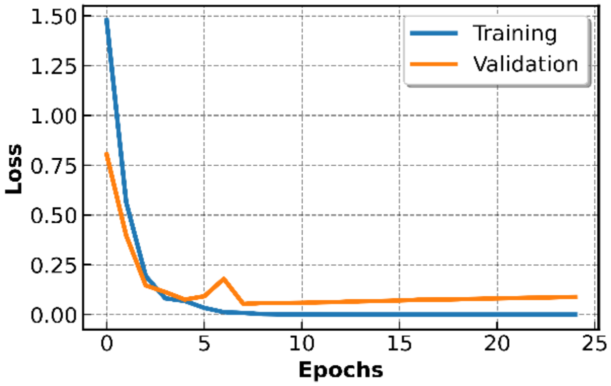


Fig. 6. Loss Curve of the Proposed Model.

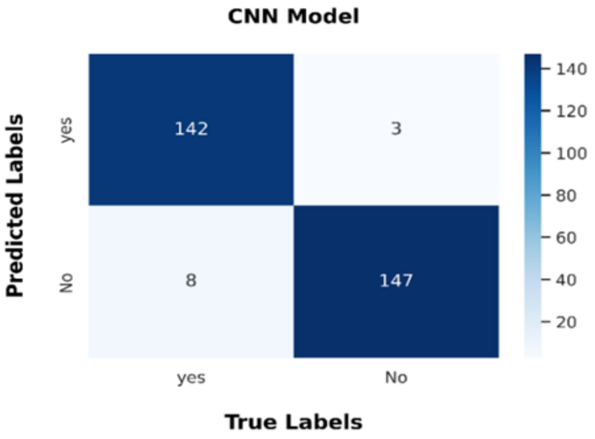


Fig. 7. Confusion Matrix of the Proposed Model.

According to the results of the values obtained by the confusion matrix in Fig. 7, it’s shown that 142 positive cases of brain tumors from 300 were correctly classified by the model and 147 Non-tumorous brain images were correctly classified as belonging to the negative class by the model. The TP, TN, FP and FN values emerged by the confusion matrix are then used by the code for the calculations of the Loss Metrics accuracy, as well as the precision values and sensitivity. The results of the evaluation metrics are provided in Table II.

Another way to visualize the sensitivity and specificity of the model is by creating a ROC curve, which is a plot that displays the sensitivity and specificity of a logistic regression model. The AUC ROC equals the probability of having a randomly chosen “true” rating greater than a randomly chosen “false” rating. In the case of an ideal classification, the AUC ROC value is equal to 1, for a value of 0.5 (case of the diagonal line on the curve) this explains that the classifier only guesses [56].The plot of the ROC curve in Fig. 8 indicates that the highest area is under the curve with a percentage of 96%, which means that the CNN model correctly classifies the images into their categories. Table III report all the performance evaluation metrics of the proposed model.

C. Discussion

Our experimental results demonstrate that the proposed CNN model converges better than the ANN approach, the Random Forest classifier, Transfer learning algorithms, and other CNN models. As shown in Table IV, the proposed model achieved the best accuracy rate of 96% and the best F1-Score of 96.5% with a precision of 98%. These values are high and well ranked compared to the results obtained by the other models already mentioned. Table IV compares the different models and allows us to conclude that our model is the best ranked in terms of accuracy.

Through the analysis of the accuracy and loss curves during the two periods of training and validation, the graph proves that the model has no overfitting issues and returns an average loss value of 0.28, which means that the model is doing well in predicting tumor health and unhealthy brains. The ROC curves show that our proposed CNN model is a reliable system for the detection and classification of brain tumors.

TABLE III. PERFORMANCE EVALUATION OF THE PROPOSED MODEL BASED ON SCORING METRICS

Evaluating Metrics	**Performance Score**
Loss	0,2896
Accuracy	96.33%
Precision	97.93%
Sensitivity	95%
F1–Score	96.44%
Specificity	75.72%

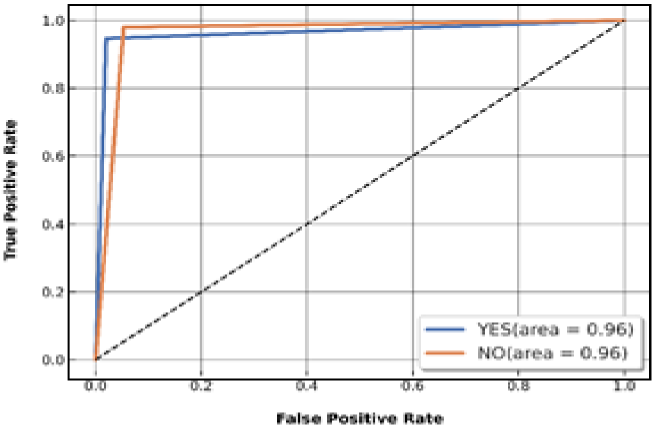


Fig. 8. ROC Curve of the Proposed Model.

In addition to the research work, the performance of the proposed model has been compared with the results obtained by other studies focused on the same case study of brain tumors. For example, the authors in [57] proposed a 7-layered 2 CNN; they obtained an accuracy of 95%. Another technique proposed in [58] using cascaded CNN for segmentation, Vgg19 and data augmentation approach, obtained an accuracy value of 86.7%, 78.9%, and 95.6% for AD, lesion, and normal class. Authors in [59] used the support vector machine and 82% accuracy was obtained. Compared to these models, our proposed CNN model remains the best ranked in terms of accuracy, whether for those adopting the same convolutional neural network architectures or those using other segmentation techniques.

TABLE IV. COMPARATIVE SUMMURY OF DIFFFERENT CLASSIFIERS FOR SEGMENTATION ACCURACY

Study	**Method/Classifier**	**Accuracy Rate**
[47]	CNN	93%
[50]	MobileNet V2	92%
	Inception V3	91%
	VGG19	88%
[51]	ANN	91,8%
[52]	Radom forest with ROI process	83,6%
	Radom forest without ROI process	87,6%
Proposed Study	**CNN Model**	**96%**

VI. Conclusion and Persepectives

In this paper, a CNN model for the segmentation of MRI images of brain tumors into two classes with tumors and without tumors is proposed. The proposed method for detection and classification of MRI images provided the best accuracy achieved by other neural network models. These medical images have undergone preprocessing and resizing before being processed by the convolutional neural network. Training and validation were performed on 3,000 high-resolution MRI images. The performance of the CNN model is evaluated using several evaluation metrics. Through this experiment, the proposed model is found to outperform other CNN models in several performance aspects, including 96% overall accuracy and 98% accuracy. Finally, for the given dataset, CNN proves to be the best technique to predict the presence of brain tumors. Based on performance evaluation metrics and curve analysis, this work demonstrates the ability of a CNN network to detect and classify tumors in the brain with a higher accuracy rate. This work has presented the architecture of convolutional neural networks and has demonstrated their performance when applied to an adjusted database of brain images. In future work, the architecture of the proposed model could be perfected, and its reliability and performance will be evaluated with a large database.

References

[1] Q. Nida-Ur-Rehman, I. Ahmed, G. Masood, N.-U.-S. -, M. Khan, and A. Adnan, “Segmentation of Brain Tumor in Multimodal MRI using Histogram Differencing & KNN,” ijacsa, vol. 8, no. 4, 2017, doi: 10.14569/IJACSA.2017.080434.

[2] S. M. Kulkarni and G. Sundari, “A Framework for Brain Tumor Segmentation and Classification using Deep Learning Algorithm,” IJACSA, vol. 11, no. 8, 2020, doi: 10.14569/IJACSA.2020.0110848.

[3] K. Ejaz et al., “Segmentation Method for Pathological Brain Tumor and Accurate Detection using MRI,” ijacsa, vol. 9, no. 8, 2018, doi: 10.14569/IJACSA.2018.090851.

[4] J. Sikder, U. K. Das, and R. J. Chakma, “Supervised Learning-based Cancer Detection,” IJACSA, vol. 12, no. 5, 2021, doi: 10.14569/IJACSA.2021.01205101.

[5] H. Moujahid, B. Cherradi, and L. Bahatti, “Convolutional Neural Networks for Multimodal Brain MRI Images Segmentation: A Comparative Study,” in Smart Applications and Data Analysis, vol. 1207, M. Hamlich, L. Bellatreche, A. Mondal, and C. Ordonez, Eds. Cham: Springer International Publishing, 2020, pp. 329–338. doi: 10.1007/978-3-030-45183-7_25.

[6] A. Mustaqeem, A. Javed, and T. Fatima, “An Efficient Brain Tumor Detection Algorithm Using Watershed & Thresholding Based Segmentation,” IJIGSP, vol. 4, no. 10, pp. 34–39, Sep. 2012, doi: 10.5815/ijigsp.2012.10.05.

[7] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, “Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks,” arXiv:1705.03820 [cs], Jun. 2017, Accessed: Mar. 31, 2022. [Online]. Available: http://arxiv.org/abs/1705.03820.

[8] O. Bouattane, B. Cherradi, M. Youssfi, and M. O. Bensalah, “Parallel c-means algorithm for image segmentation on a reconfigurable mesh computer,” Parallel Computing, vol. 37, no. 4–5, pp. 230–243, Apr. 2011, doi: 10.1016/j.parco.2011.03.001.

[9] N. A. Ali, A. E. abbassi, and B. Cherradi, “The performances of iterative type-2 fuzzy C-mean on GPU for image segmentation,” J Supercomput, Jun. 2021, doi: 10.1007/s11227-021-03928-9.

[10] N. Aitali, B. Cherradi, A. El Abbassi, O. Bouattane, and M. Youssfi, “GPU based implementation of spatial fuzzy c-means algorithm for image segmentation,” in 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), Tangier, Morocco, Oct. 2016, pp. 460–464. doi: 10.1109/CIST.2016.7805092.

[11] Jin Liu, Min Li, Jianxin Wang, Fangxiang Wu, Tianming Liu, and Yi Pan, “A survey of MRI-based brain tumor segmentation methods,” Tinshhua Sci. Technol., vol. 19, no. 6, pp. 578–595, Dec. 2014, doi: 10.1109/TST.2014.6961028.

[12] N. Ait Ali, B. Cherradi, A. El Abbassi, O. Bouattane, and M. Youssfi, “GPU fuzzy c-means algorithm implementations: performance analysis on medical image segmentation,” Multimed Tools Appl, vol. 77, no. 16, pp. 21221–21243, Aug. 2018, doi: 10.1007/s11042-017-5589-6.

[13] O. El Gannour, B. Cherradi, S. Hamida, M. Jebbari, and A. Raihani, “Screening Medical Face Mask for Coronavirus Prevention using Deep Learning and AutoML,” in 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, Mar. 2022, pp. 1–7. doi: 10.1109/IRASET52964.2022.9737903.

[14] W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, “Deep CNN for Brain Tumor Classification,” Neural Process Lett, vol. 53, no. 1, pp. 671–700, Feb. 2021, doi: 10.1007/s11063-020-10398-2.

[15] O. El Gannour et al., “Concatenation of Pre-Trained Convolutional Neural Networks for Enhanced COVID-19 Screening Using Transfer Learning Technique,” Electronics, vol. 11, no. 1, p. 103, Dec. 2021, doi: 10.3390/electronics11010103.

[16] S. Hamida, O. El Gannour, B. Cherradi, A. Raihani, H. Moujahid, and H. Ouajji, “A Novel COVID-19 Diagnosis Support System Using the Stacking Approach and Transfer Learning Technique on Chest X-Ray Images,” Journal of Healthcare Engineering, vol. 2021, pp. 1–17, Nov. 2021, doi: 10.1155/2021/9437538.

[17] O. El Gannour, S. Hamida, B. Cherradi, A. Raihani, and H. Moujahid, “Performance Evaluation of Transfer Learning Technique for Automatic Detection of Patients with COVID-19 on X-Ray Images,” in 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, Dec. 2020, pp. 1–6. doi: 10.1109/ICECOCS50124.2020.9314458.

[18] S. Hamida, O. E. Gannour, B. Cherradi, H. Ouajji, and A. Raihani, “Optimization of Machine Learning Algorithms Hyper-Parameters for Improving the Prediction of Patients Infected with COVID-19,” in 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, Dec. 2020, pp. 1–6. doi: 10.1109/ICECOCS50124.2020.9314373.

[19] O. El Gannour, S. Hamida, S. Saleh, Y. Lamalem, B. Cherradi, and A. Raihani, “COVID-19 Detection on X-Ray Images using a Combining Mechanism of Pre-trained CNNs,” IJACSA, vol. 13, no. 6, 2022, doi: 10.14569/IJACSA.2022.0130668.

[20] O. Asmae, R. Abdelhadi, C. Bouchaib, S. Sara, and K. Tajeddine, “Parkinson’s Disease Identification using KNN and ANN Algorithms based on Voice Disorder,” in 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, Apr. 2020, pp. 1–6. doi: 10.1109/IRASET48871.2020.9092228.

[21] S. Laghmati, B. Cherradi, A. Tmiri, O. Daanouni, and S. Hamida, “Classification of Patients with Breast Cancer using Neighbourhood Component Analysis and Supervised Machine Learning Techniques,” in 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, Sep. 2020, pp. 1–6. doi: 10.1109/CommNet49926.2020.9199633.

[22] O. Daanouni, B. Cherradi, and A. Tmiri, “Predicting diabetes diseases using mixed data and supervised machine learning algorithms,” in Proceedings of the 4th International Conference on Smart City Applications, Casablanca Morocco, Oct. 2019, pp. 1–6. doi: 10.1145/3368756.3369072.

[23] H. Moujahid, B. Cherradi, and L. Bahatti, “Comparison Study on Some Convolutional Neural Networks for Cerebral MRI Images Segmentation,” in Distributed Sensing and Intelligent Systems, M. Elhoseny, X. Yuan, and S. Krit, Eds. Cham: Springer International Publishing, 2022, pp. 557–568. doi: 10.1007/978-3-030-64258-7_48.

[24] O. Terrada, S. Hamida, B. Cherradi, A. Raihani, and O. Bouattane, “Supervised Machine Learning Based Medical Diagnosis Support System for Prediction of Patients with Heart Disease,” Adv. sci. technol. eng. syst. j., vol. 5, no. 5, pp. 269–277, 2020, doi: 10.25046/aj050533.

[25] O. Terrada, A. Raihani, O. Bouattane, and B. Cherradi, “Fuzzy cardiovascular diagnosis system using clinical data,” in 2018 4th

International Conference on Optimization and Applications (ICOA), Mohammedia, Apr. 2018, pp. 1–4. doi: 10.1109/ICOA.2018.8370549.

[26] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, “Atherosclerosis disease prediction using Supervised Machine Learning Techniques,” in 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, Apr. 2020, pp. 1–5. doi: 10.1109/IRASET48871.2020.9092082.

[27] B. Cherradi, O. Terrada, A. Ouhmida, S. Hamida, A. Raihani, and O. Bouattane, “Computer-Aided Diagnosis System for Early Prediction of Atherosclerosis using Machine Learning and K-fold cross-validation,” in 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, Jul. 2021, pp. 1–9. doi: 10.1109/ICOTEN52080.2021.9493524.

[28] S. Hamida, B. Cherradi, O. El Gannour, O. Terrada, A. Raihani, and H. Ouajji, “New Database of French Computer Science Words Handwritten Vocabulary,” in 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, Jul. 2021, pp. 1–5. doi: 10.1109/ICOTEN52080.2021.9493438.

[29] S. Hamida, B. Cherradi, H. Ouajji, and A. Raihani, “Convolutional Neural Network Architecture for Offline Handwritten Characters Recognition,” in Innovation in Information Systems and Technologies to Support Learning Research, vol. 7, M. Serrhini, C. Silva, and S. Aljahdali, Eds. Cham: Springer International Publishing, 2020, pp. 368–377. doi: 10.1007/978-3-030-36778-7_41.

[30] S. Hamida, B. Cherradi, and H. Ouajji, “Handwritten Arabic Words Recognition System Based on HOG and Gabor Filter Descriptors,” in 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, Apr. 2020, pp. 1–4. doi: 10.1109/IRASET48871.2020.9092067.

[31] S. Hamida, B. Cherradi, O. Terrada, A. Raihani, H. Ouajji, and S. Laghmati, “A Novel Feature Extraction System for Cursive Word Vocabulary Recognition using Local Features Descriptors and Gabor Filter,” in 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, Sep. 2020, pp. 1–7. doi: 10.1109/CommNet49926.2020. 9199642.

[32] M. Jebbari, B. Cherradi, O. El Gannour, S. Hamida, and A. Raihani, “Exploration Study on Learning Styles Identification and Prediction Techniques,” in 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, Mar. 2022, pp. 1–7. doi: 10.1109/IRASET52964. 2022.9738030.

[33] L. Ajallouda, K. Najmani, A. Zellou, and E. habib Benlahmar, “Doc2Vec, SBERT, InferSent, and USE Which embedding technique for noun phrases?,” in 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, Mar. 2022, pp. 1–5. doi: 10.1109/IRASET52964. 2022.9738300.

[34] L. Ajallouda, O. Hourrane, A. Zellou, and E. H. Benlahmar, “Toward a New Process for Candidate Key-Phrases Extraction,” in Digital Technologies and Applications, vol. 455, S. Motahhir and B. Bossoufi, Eds. Cham: Springer International Publishing, 2022, pp. 466–474. doi: 10.1007/978-3-031-02447-4_48.

[35] L. Ajallouda, F. Z. Fagroud, A. Zellou, and E. B. Lahmar, “KP-USE: An Unsupervised Approach for Key-Phrases Extraction from Documents,” IJACSA, vol. 13, no. 4, 2022, doi: 10.14569/IJACSA.2022.0130433.

[36] N. Aitali, B. Cherradi, A. El, O. Bouattane, and M. Youssfi, “Parallel Implementation of Bias Field Correction Fuzzy C-Means Algorithm for Image Segmentation,” ijacsa, vol. 7, no. 3, 2016, doi: 10.14569/IJACSA. 2016.070352.

[37] N. A. Ali, B. Cherradi, A. El Abbassi, O. Bouattane, and M. Youssfi, “New parallel hybrid implementation of bias correction fuzzy C-means algorithm,” in 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Fez, May 2017, pp. 1–6. doi: 10.1109/ATSIP.2017.8075519.

[38] J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, “Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study,” PLOS Medicine, vol. 15, no. 11, p. e1002683, Nov. 2018, doi: 10.1371/journal.pmed.1002683.

[39] N. Bien et al., “Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet,” PLoS Med, vol. 15, no. 11, p. e1002699, Nov. 2018, doi: 10.1371/journal.pmed.1002699.

[40] X. Liu, L. Song, S. Liu, and Y. Zhang, “A Review of Deep-Learning-Based Medical Image Segmentation Methods,” Sustainability, vol. 13, no. 3, p. 1224, Jan. 2021, doi: 10.3390/su13031224.

[41] Z. Liu et al., “Deep learning based brain tumor segmentation: a survey,” Complex Intell. Syst., Jul. 2022, doi: 10.1007/s40747-022-00815-5.

[42] D. C. Febrianto, I. Soesanti, and H. A. Nugroho, “Convolutional Neural Network for Brain Tumor Detection,” IOP Conf. Ser.: Mater. Sci. Eng., vol. 771, no. 1, p. 012031, Mar. 2020, doi: 10.1088/1757-899X/771/1/012031.

[43] S. Dabeer, M. M. Khan, and S. Islam, “Cancer diagnosis in histopathological image: CNN based approach,” Informatics in Medicine Unlocked, vol. 16, p. 100231, 2019, doi: 10.1016/j.imu. 2019.100231.

[44] M. Arif, F. Ajesh, S. Shamsudheen, O. Geman, D. Izdrui, and D. Vicoveanu, “Brain Tumor Detection and Classification by MRI Using Biologically Inspired Orthogonal Wavelet Transform and Deep Learning Techniques,” Journal of Healthcare Engineering, vol. 2022, pp. 1–18, Jan. 2022, doi: 10.1155/2022/2693621.

[45] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” Insights Imaging, vol. 9, no. 4, pp. 611–629, Aug. 2018, doi: 10.1007/s13244-018-0639-9.

[46] E. Irmak, “Multi-Classification of Brain Tumor MRI Images Using Deep Convolutional Neural Network with Fully Optimized Framework,” Iran J Sci Technol Trans Electr Eng, vol. 45, no. 3, pp. 1015–1036, Sep. 2021, doi: 10.1007/s40998-021-00426-9.

[47] D. C. Febrianto, I. Soesanti, and H. A. Nugroho, “Convolutional Neural Network for Brain Tumor Detection,” IOP Conf. Ser.: Mater. Sci. Eng., vol. 771, no. 1, p. 012031, Mar. 2020, doi: 10.1088/1757-899X/771/1/012031.

[48] W. Koehrsen, “Transfer Learning with Convolutional Neural Networks in PyTorch,” Medium, Nov. 26, 2018. https://towardsdatascience. com/transfer-learning-with-convolutional-neural-networks-in-pytorch-dd09190245ce (accessed Jun. 17, 2022).

[49] C. Srinivas et al., “Deep Transfer Learning Approaches in Performance Analysis of Brain Tumor Classification Using MRI Images,” Journal of Healthcare Engineering, vol. 2022, pp. 1–17, Mar. 2022, doi: 10.1155/2022/3264367.

[50] T. Tazin et al., “A Robust and Novel Approach for Brain Tumor Classification Using Convolutional Neural Network,” Comput Intell Neurosci, vol. 2021, p. 2392395, Dec. 2021, doi: 10.1155/2021/2392395.

[51] M. Nazir, F. Wahid, and S. Ali Khan, “A simple and intelligent approach for brain MRI classification,” Journal of Intelligent & Fuzzy Systems, vol. 28, no. 3, pp. 1127–1135, 2015, doi: 10.3233/IFS-141396.

[52] I. Soesanti, M. H. Avizenna, and I. Ardiyanto, “Classification of Brain Tumor MRI Image using Random Forest Algorithm and Multilayers Perceptron,” p. 6, 2020.

[53] H. Ali Khan et al., “Brain tumor classification in MRI image using convolutional neural network,” Mathematical Biosciences and Engineering, vol. 17, no. 5, pp. 6203–6216, 2020, doi: 10.3934/mbe.2020328.

[54] “Brain_Tumor_Detection_MRI.” https://www.kaggle.com/datasets/abhranta/brain-tumor-detection-mri (accessed Jul. 01, 2022).

[55] D. R. Sarvamangala and R. V. Kulkarni, “Convolutional neural networks in medical image understanding: a survey,” Evol. Intel., vol. 15, no. 1, pp. 1–22, Mar. 2022, doi: 10.1007/s12065-020-00540-3.

[56] A. E. Maxwell, T. A. Warner, and L. A. Guillén, “Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review,” Remote Sensing, vol. 13, no. 13, p. 2450, Jun. 2021, doi: 10.3390/rs13132450.

[57] J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes, “Big data analysis for brain tumor detection: Deep convolutional neural networks,” Future Generation Computer Systems, vol. 87, pp. 290–297, Oct. 2018, doi: 10.1016/j.future.2018.04.065.

[58] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," Journal of Computational Science, vol. 30, pp. 174–182, Jan. 2019, doi: 10.1016/j.jocs.2018.12.003.

[59] N. Vani, A. Sowmya, and N. Jayamma, "Brain Tumor Classification using Support Vector Machine," vol. 04, no. 07, p. 6.

Firefly Algorithm with Mini Batch K-Means Entropy Measure for Clustering Heterogeneous Categorical Timber Data

Nurshazwani Muhamad Mahfuz[1]
Faculty of Computer and Mathematical Science
Universiti Teknologi MARA
Shah Alam, Malaysia

Marina Yusoff [2]
Institute for Big Data Analytics and Artificial Intelligence
Faculty of Computer and Mathematical Science
Universiti Teknologi MARA, Shah Alam, Malaysia

Muhammad Shaiful Nordin[3]
Malaysian Timber Industry Board
Menara PGRM, Jalan Pudu Ulu, 56100 Cheras
Kuala Lumpur, Malaysia

Zakiah Ahmad[4]
College of Engineering
Universiti Teknologi MARA
Shah Alam, Malaysia

***Abstract*—Clustering analysis is the process of identifying similar patterns in various types of data. Heterogeneous categorical data consists of data on ordinal, nominal, binary, and Likert scales. The clustering solution for heterogeneous data clustering remains difficult due to partitioning complex and dissimilarity features. It is necessary to find a solution to high-quality clustering techniques to efficiently determine the significant features of the data. This paper emphasizes using the firefly algorithm to reduce the distance gap between features and improve clustering performance. To obtain an optimal global solution for clustering, we proposed a hybrid of mini-batch k-means (MBK) clustering-based entropy distance measures (EM) with a firefly optimization algorithm (FA). This study compares the performance of hybrid K-Means, Agglomerative, DBSCAN, and Affinity clustering models with EM and FA. The evaluation uses a variety of data from the timber perception survey dataset. In terms of performance, the proposed MBK+EM+FA has superior and most effective clustering. It achieves a higher accuracy of 96.3 percent, a 97 percent F-measure, a 98 percent precision, and a 97 percent recall. Other external assessments revealed that the Homogeneity (HOMO) is 79.14 percent, the Fowlkes-Mallows Index (FMI) is 93.07 percent, the Completeness (COMP) is 78.04 percent, and the V-Measure (VM) is 78.58 percent. Both proposed MBK+EM+FA and MBK+EM took about 0.45s and 0.35s to compute, respectively. The excellent quality of the clustering results does not justify such time constraints. Surprisingly, the proposed model reduced the distance measure of all heterogeneous features. The future model could put heterogeneous categorical data from a different domain to the test.**

Keywords—Clustering; mini batch k-means; entropy; heterogeneous categorical; firefly optimization algorithm

I. Introduction

Categorical data or known as qualitative data is a type of data that can be stored and identified using the names or labels that have been assigned to it. Most statistical analysis approaches are incompatible with it, and only bar graphs and pie charts can visualize the data. Due to the rapid emergence and growth of information, it has become increasingly important to discover the group structure of objects within them. It also has difficulty extracting helpful information. One of the most effective methods that can extract such information is clustering. This technique allows organizing the data to access the information. Using clustering techniques, data analysts can easily extract valuable information from large datasets without supervision. The categorical clustering techniques were widely used in many real-world applications such as security analytics [1][2] and solving cold-start recommendation problems [3].

The traditional methods commonly used for data clustering problems are hierarchical and partitional. In some cases, the clustering process relies on distance or similarity measures. In a clustering algorithm, the data objects are typically represented in Euclidean distance in *k*-means. The clustering process's objective is to minimize the square distance from the cluster center to the cluster domain. *k*-means is widely used as a clustering algorithm and is effective when dealing with enormous volumes of data. However, *k*-means cannot be directly applied to data sets with categorical features. The transformation and parameter adjustment into numerical form is required since machines cannot interpret the categorical features directly. Label encoding, one hot encoding, and dummy variable encoding are methods for converting categorical data into numerical data.

Meanwhile, *k*-modes using simple matching dissimilarity measures can directly be applied for purely categorical data clustering. As categorical data cannot be estimated using mean or medians, the Euclidean distance metric was replaced with a simple matching dissimilarity measure, and the mean calculation for representing centroids was substituted with modes. However, these methods have some disadvantages, such as empty groups may appear in the first step of the solution and the final division of data is not optimal due to the appearance of extreme points. It is also trapped in local optima and local maxima and is sensitive to initial cluster centers.

Nature-inspired algorithms have gained much attention as global optimization tools to assist in solving various real-life complex optimization problems such as profit production [4], scheduling [5], queuing system [6], market segmentation, and opinion mining [7]. Nature-inspired algorithms have received a lot of attention as global optimization tools to help with real-life complex optimization problems like profit production [4], scheduling [5], queuing systems [6], market segmentation, and opinion mining [7].

Stochastic methods can be used to overcome clustering problems. Optimization methods refer to finding feasible solutions for problems to give an efficient, robust solution. Recent research has used Artificial Bee Colony (ABC) [8] and Cuckoo Search [9] to improve categorical data clustering quality. Optimizing a complex clustering problem is difficult due to the lack of a single measure that works best for heterogeneous categorical data.

This paper proposes the hybrid firefly algorithm and MBK to improve clustering performance using entropy distance metrics as similarity measures. The experimental results were also compared using the entropy distance method with Mini Batch *k*-Means (MBK), *k*-Means, Agglomerative Hierarchical, Density-based spatial clustering of applications with noise (DBSCAN), and Affinity Propagation clustering methods on a public survey data.

The rest of the paper follows the organization of the section as follows. Section II describes the related works. Section III explains the firefly algorithm. The mini-batch k-means algorithm is in Section IV. The proposed algorithm is introduced in Section V. The experimental results are discussed in Section VI. Section VII and Section VIII are the discussion and conclusion of the paper.

II. Related Work

The related work on categorical clustering, FA, and clustering models that use metaheuristic methods are reviewed in this section. Nominal, binary, ordinal, and Likert data types are categorical [10]. The combination of these data types is simultaneously considered to have heterogeneous information or data. The main goal of data clustering is to predict and find the groups for each data object from unlabeled data. On the other hand, selecting appropriate representations for data is one of the central problems in machine learning and data mining.

Clustering categorical data is very challenging. The challenge includes processing non-categorical variables and a required procedure for applying the similarity measures for the matching process. Regarding the grouping data from previous studies, the set-valued *k*-modes algorithm outperforms three existing categorical clustering techniques and is scalable to big datasets [11]. Algorithm new *k*-means like a method for categorical clustering data has shown that the proposed clustering method outperformed the k-means [12]. A Unified Entropy-Based Distance Metric for ordinal-and-nominal attributes on both real and benchmark data sets shows that the proposed metric surpasses the existing alternatives. A Unified Entropy-Based Distance Metric for ordinal-and-nominal attributes on both real and benchmark data sets shows that the proposed metric surpasses the existing alternatives [13]. A holo-entropy-based hierarchical clustering technique for categorical data outperforms other known algorithms in terms of efficiency, accuracy, and reproducibility [14].

The linear programming model outperformed the traditional and other enhanced *k*-modes algorithms on categorical datasets [8]. Learning-Based dissimilarity for categorical data clustering outperforms in terms of several performance indicators [16]. Compared to *k*-modes, the *k*-Approximate Modal Haplotype achieves an average performance increase of 0.51 percent in Precision and 0.40 percent in Normalized Discounted Cumulative Gain. However, because of their scalability, *k*-modes are more flexible to utilize [3]. The conventional categorical clustering performed well but provided local optimum solutions that affect data division.

The firefly optimization algorithm is one of the techniques that has been effectively used to solve issues in several areas due to its global optimum solution, resilience, efficiency, and capacity to handle problems in various sectors, including NP-hard, versatility, and other outstanding benefits. A comprehensive overview of FA that covers the various domains where the method is applied to a wide range of real-world applications with satisfying clustering results. Regarding clustering validity metrics, the FA shows that the new clustering methods outperform existing clustering and other hybrid metaheuristic methodologies [17]. In both distance and performance measurements for clustering tasks, the inward intensified exploration fa and compound intensified exploration fa models show statistically significant superiority [18]. Both algorithms are proposed to overcome the limitation of the FA model and *k*-means clustering. Compared to other algorithms, the firefly algorithm with *k*-means clustering produces better results, demonstrating the usefulness of the firefly algorithm with *k*-means clustering in offering a competitive solution to the traveling salesman problem [19]. *k*-means clustering with modifying the firefly algorithm is significantly more efficient than other algorithms [20]. From this, it can be seen that the combination of FA and conventional clustering algorithms outperform them in many real-world data sets, including numerical data.

Furthermore, in the previous literature, some clustering techniques for categorical datasets combined with other hybridization of optimization have been investigated and outperform the traditional *k*-modes algorithm in several aspects [21], as well as Fuzzy *k*-partition based on the ABC, outperforms the baseline algorithm in terms of its validity clustering for categorical data clustering [8]. The ABC algorithm was inspired by the foraging habits of bees and is one of the swarm-based metaheuristic optimization algorithms. Several studies in categorical clustering for hybridization of global optimization using a conventional clustering algorithm. However, few studies on hybridizing global optimization algorithms and conventional clustering algorithms use entropy as a distance measure.

Therefore, in this research, MBK+EM+FA has proposed to find the optimum global result for heterogeneous categorical data clustering using entropy distance as a similarity measure.

Because of its automatic subdivision and capacity to deal with multimodality, FA is preferred above other algorithms [22].

III. Firefly Algorithm (FA)

Recently flashing behavior of fireflies has been identified as unique to the species. FA is established by Xin She Yang [15], and it was based on the idealized behavior of the flashing characteristics of fireflies. From the literature, it is found that the FA algorithm can outperform when compared to many other algorithms. FA algorithm expands, and new variants emerge to solve all optimization problems. FA is chosen instead of other algorithms due to its simple, flexible, fast convergence, which efficiently solves many real-world problems.

FA is a swarm-intelligence-based algorithm, so it has similar advantages to the other swarm intelligence-based algorithms. FA has two significant advantages over other algorithms: automatic subdivision and the ability to deal with multimodality. FA offers simplicity, flexibility, and ease of implementation. Recently, FA is one of the bio-inspired algorithms used to solve clustering problems inspired by biochemical and social aspects of real fireflies [19]. Fireflies' behaviors, such as short and rhythmic flashes, can be considered operators of computational intelligence methods. The basic assumptions formulation of the FA algorithm is as follows.

- The intensity of a firefly decreases with the increase in distance. The firefly attracts the firefly that is closer to it. The light intensity I decrease as the distance r^2 increases, $I \propto \frac{1}{r^2}$.
- All fireflies are unisex and attracted to other fireflies regardless of their sex.
- The objective function defines the brightness of a firefly.

The FA algorithm has two critical features: variation in the light intensity and formulation of attractiveness (Yang 2010). The flashes are used as a communication tool by fireflies. Light is absorbed in the media, so the attractiveness of two fireflies will vary with the degree of absorption. The light intensity, $I\ (r)$ varies to the inverse square law as stated in equation 1.

$$I\ (r) = \frac{I_s}{r^2} \tag{1}$$

The light intensity, I, and the absorption coefficient, γ, varies with the distance, r for a given as in equation 2.

$$I\ = I_0 e^{-\gamma r} \tag{2}$$

Where I_0 is the original of the light intensity. The combined effect of inverse square law and absorption can be assumed as Gaussian form and represented in equation 3.

$$I\ = I_0 e^{-\gamma r^2} \tag{3}$$

The attractiveness of fireflies is proportional to the light intensity of the adjacent fireflies. The attractiveness, β of firefly is in equation 4.

$$\beta(r) = \beta_0 e^{-\gamma r^2} \tag{4}$$

Where β_0 is attractiveness at $r = 0$, the Euclidean distance between the firefly, i at x_i, and firefly, j at x_j, is given by equation 5.

$$r_{ij} = \sqrt{\sum_{k=1}^{d}(x_{ik} - x_{jk})^{\ 2}} \tag{5}$$

The computation movement of firefly i is attracted to another more attractive, brighter firefly j, and the formula is in equation 6.

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2}(x_i - x_j) + \alpha\left(rand - \frac{1}{2}\right) \tag{6}$$

Where α is the randomization parameter and $rand$ is a random number generator uniformly distributed between 0 and 1. The second represents the attraction, and the third term is randomization.

Algorithm 1: The Pseudocode of FA

```
Start
Initialize algorithm parameters:
Define the objective function $f(x)$, where $x = (x_1, \ldots\ldots, x_d)$
Generate the initial population of fireflies $x_i$ $(i = 1, 2, \ldots\ldots, n)$
Determine the light intensity of the firefly at $x_i$ by using objective function $f(x_i)$
**While** ($K$<Maximum_Generation) // Where $k = 1$ to maximum
    **for** $p = 1 : n$ // all $n$ fireflies
        **for** $q = 1$ to $n$ //$n$ fireflies
            if ($I_p$< $I_q$)
                Move firefly $p$ towards $q$ by using
                $x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2}(x_i - x_j) + \alpha\left(rand - \frac{1}{2}\right)$
            **end** if
            Calculate new solutions and update light intensity for the next iteration
        **end** for $q$
    **end** for $p$
    Sort the fireflies based on the intensity value and find the current best solution.
end while
```

IV. Mini Batch K-Means (MBK) Clustering

A distributed random batching strategy known as Mini Batch K-means was used to store and update the data incrementally. Data and prototype values from each batch were used to update the cluster—the learning rate increases with the number of iterations in a batch. Before reaching a consensus, clusters must go through several iterations. There are several benefits to using MBK, such as its shorter computation time, the most simple unsupervised learning that solves clustering methodologies, and higher accuracy when dealing with mixed and huge datasets [23].

MBK+EM (Mini Batch K-Means with Entropy Measure) is an embedded entropy distance measure with Mini Batch K-Means that aims to determine the quality of the performance of clustering in heterogeneous categorical data. The previous experiment demonstrates that the Mini Batch K-mean with Entropy Measure at k =2 outperformed other clustering algorithms in clustering accuracy at 88%, V-Measure at 0.82, Adjusted Rand Index at 0.87, and Fowlkes-Mallow's Index at 0.94. The experiment was fixed seven times the average minimum elapsed time-varying for cluster generation, k at 0.26s.

V. Proposed Algorithm

The MBK algorithm has an advantage in reducing the amount of computation to converge to a local solution [24]. However, it has a drawback to finding the local optimum clustering results and the existing FA, which has a problem remembering the best solution for each firefly in the past. When there are no brighter-colored fireflies around, it also moves at random. The FA in this research performs the clustering procedure with the optimal centroid point. Therefore, the MBK is combined with the firefly optimization algorithm to obtain the global optimum solution and gives the firefly method a substantial improvement in clustering performance.

FA was chosen over other optimization algorithms because previous research shows that hybridizing FA with conventional clustering algorithms improves result quality [19]. The MBK+EM+FA algorithm combines the advantages of its distance measure, which would be possible to improve the clustering performance. One step of the MBK algorithm is utilized at the end of all the iterations of FA.

Fig. 1 depicts the process flow of the proposed hybridization FA with MBK+EM. The MBK+EM operator is incorporated into the FA to locate the centroid. The fireflies distribute randomly in the search space based on the objective function from the population at random from the given data objects. The distance between the firefly's position and the actual data in the dataset determines the intensity of each firefly. After analyzing the distance, all data shows the minimum distance value among the fireflies. The movement of the fireflies in the search space indicates the firefly's brightness. The iterative process of the swarm involves a comparison of the intensity of one firefly to another firefly, and the firefly's brightness defines its firefly movement. The attractiveness varies according to the distance between the two fireflies. Then, calculate the intensity for new fireflies based on firefly movement. To determine the new position, apply the MBK+EM operator to the entire population of fireflies. For each firefly, the MBK+EM operator is used to compute the mean value of all associated objects. Then, update the intensity values and evaluate the new firefly position for the entire firefly. Before proceeding to the next iteration, the fireflies' selection depends on their intensity value. The result is continuously updated during iterative processes until the stopping criteria are met. A post-process to select the final centroids for determining the best solution while ranking the position of the fireflies.

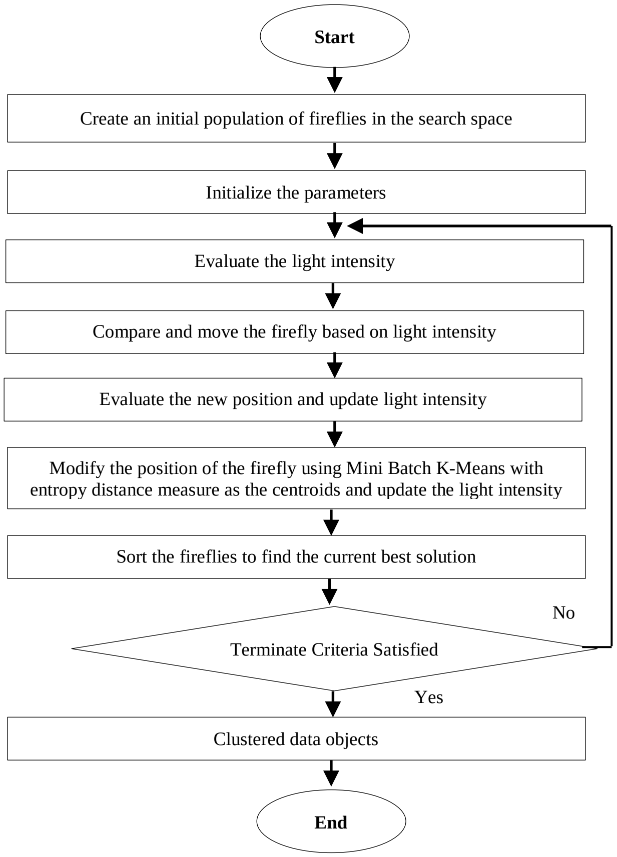

Start
Create an initial population of fireflies in the search space
Initialize the parameters
Evaluate the light intensity
Compare and move the firefly based on light intensity
Evaluate the new position and update light intensity
Modify the position of the firefly using Mini Batch K-Means with entropy distance measure as the centroids and update the light intensity
Sort the fireflies to find the current best solution
No
Terminate Criteria Satisfied
Yes
Clustered data objects
End

Fig. 1. Flowchart of Proposed MBK+EM+FA Algorithm.

The proposed Hybrid FA+MBK algorithm with EM phase is indicated in Algorithm 2. The following is a brief overview of all aspects of the Hybrid FA algorithm. The proposed algorithm starts by initializing the algorithm parameter in Step 2. The declaration of objective function performs in Step 3. The next step is randomly initiating fireflies' initialization. The position of the firefly represents the centroid of the clustering problem. Step 5 aims to determine the light intensity of the firefly to calculate the distance between the position of the fireflies. It is the initialization phase to estimate the light intensity of each firefly. Step 6-Step 10 demonstrates the movement of the fireflies in the search space, indicating the firefly's brightness. The intensity of one firefly compared with other fireflies during the iterative process in the swarm and the firefly's brightness defines its firefly movement.

The intensity value estimates the new position and place of the firefly obtained after the completion of movement calculation in steps 11 and 12. Global optima were used in the proposed firefly algorithm to control the movement of the firefly. This research will update a maximum global optimum in any iteration of the algorithm. Based on fireflies' brightness, fewer fireflies will move towards the brightest firefly. Then, the light intensity will be updated, and the current feasible or optimal solution will be found.

In steps 13 and 14, the MBK with EM is applied to the entire population of the fireflies to find a new position by updating light intensity. In the proposed FA, entropy distance was used to compute the distance of fireflies to global optima.

Step 15 focuses on the optimal values in each cluster that could discover after all the data are clustered by sorting the light intensity. Finally, the iteration executes until the maximum number of iterations.

Algorithm 2: The Pseudocode of MBK+EM+FA

```
1.  Input
2.  Initialize algorithm parameters
3.  Define the objective function $f(x)$ , where $x = (x_1, \ldots\ldots, x_d)$
4.  Generate the initial population of fireflies $x_i$ $(i = 1, 2, \ldots\ldots, n)$
5.  Determine the light intensity of the firefly at $x_i$ by using objective function $f(x_i)$
6.  While ( $K$ <Maximum_Generation) // Where $k = 1$ to maximum
7.      for $p = 1 : n$ // all $n$ fireflies
8.          for $q = 1$ to $n$ //$n$ fireflies
9.              if ($I_p$< $I_q$)
                Move firefly $p$ towards $q$ by using
                $x_i' = x_i + \beta_0 e^{-\gamma r_{ij}^2} + \propto \left(rand - \frac{1}{2}\right)$
10.             end if
11.             Calculate new solutions and update light intensity
12.         end for $q$
13.         Apply MBK with entropy distance measure, then find new solutions and update light intensity
14.     end for $p$
15.     Sort the fireflies based on the intensity value and find the current best solution;
16. end while
17. Output: Clustered data objects
```

VI. Experimental Results

A. Dataset

The experiment was performed using the secondary data of a survey on public timber utilization. The data pre-processing and cleaning procedures included removing unwanted observations, fixing the data structure, and imputing the missing data. The number of instances is 2407 was obtained from the Malaysian Timber Industry Board (MTIB), Malaysia. The dataset consists of 111 categorical features such as race, type of housing, level of knowledge, etc. This type of data is considered heterogeneous categorical data [25].

B. Parameter Settings

This subsection explains the parameter, notation, and value associated with FA evaluation. We adapt the same parameter values as suggested by the originator of FA in the late year between 2007 and 2008 [26]. The population size is 2407, the total number of respondents involved. As a stopping criterion, we also set the maximum number of iterations equal to 25. Rosenbrock is used as a benchmark objective function since previous research has shown that Rosenbrock is one of the objective functions that has shown successful performance [27]. This research implemented a static parameter for the generation and number of fireflies. The maximum value for Beta (β), Alpha (α), and Gamma (γ) is 1, 0.2, and, respectively. The algorithm parameters are summarized in Table I.

TABLE I. PARAMETER SETTING

Parameter	Notation of Parameter	Parameter Value
Brightness	Objective Function	Rosenbrock
Beta (β)	Attractiveness	1
Alpha (α)	Randomization of Parameter	0.2
Gamma (γ)	Light Absorption Coefficient	1
Number of Fireflies (n)	Population	2407
Number of Generations (g)	Iteration	25

C. Performance Measures

An external validation measure and a confusion matrix and used for the performance measure. The external validation measure is Homogeneity (HOMO), Fowlkes-Mallows Index (FMI), Completeness (COMP), and V-Measure (VM). The validation assures how good the clustering solutions are by their different ways of computations. This measure aims to measure and validate the clustering quality [28].

In the confusion matrix [16], there were four performance metrics such as true positive (TP), false positive (FP), true negative (TN), and false-negative (FN). TP is the metric that could accurately predict the optimized feature from the features in feature space collection. In contrast, the TN metric predicts the weaker feature or incorrect feature relevant to the diabetes classification process in feature space, and the FP measure can predict the incorrect diabetes feature in feature space. FN rate is an outcome of the model incorrectly predicting the negative feature effectively.

F-measure (F) is a combination of precision and recall that measures the cluster that contains only objects of a particular class and is used to balance false negatives by weighting recall parameter $\eta \geq 0$. The formula of the F-measure is in Equation 7.

$$F = \frac{(\eta^2 + 1)\ P \times R}{\eta^2 \times P + R} \tag{7}$$

Precision (P) estimates the ratio of the true positives among the cluster. The formula of Precision is in Equation 8.

$$P = \frac{TP}{TP + FP} \tag{8}$$

Recall (R) is a combination of all objects grouped into a specific class. The formula of recall is in Equation 9.

$$R = \frac{TP}{TP + FN} \tag{9}$$

Fowlkes-Mallows Index (FMI) quantifies the performance of a clustering technique by comparing it to other clusters. A score close to zero indicates largely independent labeling, whereas a value close to one reflects clustering agreement. The formula for FMI is in Equation 10.

$$FMI = \frac{TP}{\sqrt{(TP + FP)(TP + FN)}} \tag{10}$$

Accuracy (ACC) is defined only as the proportion of the actual results. The accuracy measure can be referred to in Equation 11.

$$ACC = \frac{1}{n}\sum_{i=1}^{k} a_i \qquad (11)$$

Where a_i is the number of data objects in both clusters, i and k are the numbers of clusters, and n is the total number of objects in the dataset. HOMO covers all clusters that contain only data points members of a single class. A score between 0.0 and 1.0 is obtained. A score of 1.0 stands for perfectly homogeneous labeling.

$$HOMO = -\sum_{c,k}^{1} \frac{n_{ck}}{N} \, log \, log \, (\frac{n_{ck}}{n_k}) \qquad (12)$$

Completeness (COMP) is considered comprehensive if it incorporates all data points that belong to a given class. A score between 0.0 and 1.0 is obtained. A labeling score of 1.0 indicates perfect labeling. V-measure can be used to ascertain the degree of agreement between two clustered datasets that have been clustered independently. The formula of completeness defined in Equation 13.

$$COMP = 1 - \sum_{c,k}^{1} \frac{n_{ck}}{N} \, log \, log \, (\frac{n_{ck}}{n_k}) \qquad (13)$$

V-Measure (VM) is the harmonic average between homogeneity and completeness. It can be used to determine the degree of agreement between two clustered datasets that have been clustered independently. Furthermore, if any of the two VM failed to meet the criteria, the clustering number remains zero.

$$V - Measure\ (VM) = 2 \times \frac{(HOMO \times COMP)}{(HOMO + COMP)} \qquad (14)$$

D. Experimental Results

It has been explained earlier than due to local optimum issue which aimed to improve performance of clustering external validity using heterogeneous categorical data. Several clustering algorithms were accelerated and tested with the coincident clustering problem to produce good clustering results. Moreover, the categorical nature of data creates additional complexity in clustering. MBK + EM + FA has been proposed to address such limitations of existing clustering algorithms. The technique's efficiency can be evaluated by measuring the quality of clustering results of various parameters. In order to judge the performance of the proposed technique over state-of-the-art algorithms of the other four hybrid clustering algorithms, such as a hybrid of *k*-means, Agglomerative, DBSCAN, and Affinity with EM distance metric, we conduct several experiments. Thus, this section highlights the experimental results for all five clustering algorithms: MBK, *k*-means, Agglomerative, DBSCAN, and Affinity embedded with FA and without FA using the performance measures and the computational time.

Table II demonstrates the performance of the algorithms mainly on f-measure, Precision, and recall. To add, it is interesting to note that the proposed MBK+EM+FA outperforms other algorithms with the highest in terms of F-measure, Precision, and recall, with 97%, 98%, 97%, and 96.30%, respectively. It has increased more than 0.15 of F-measure, 12% of Precision, and 16% of recall compared to MBK+EM, which is no FA is embedded as an optimization strategy. The performance of the proposed clustering algorithm was also compared with other clustering algorithms.

TABLE II. CLUSTERING PERFORMANCE FOR F-MEASURE, PRECISION AND RECALL

Clustering Algorithms	**F**	**P**	**R**
MBK+EM+FA	0.97	0.98	0.97
MBK+EM	0.82	0.86	0.81
K-Means+EM+FA	0.44	0.36	0.60
K-Means+EM	0.16	0.19	0.14
Agglomerative+EM+FA	0.44	0.36	0.60
Agglomerative+EM	0.16	0.19	0.14
DBSCAN+EM +FA	0.45	0.36	0.59
DBSCAN+EM	0.17	0.12	0.32
Affinity+EM+ FA	0.02	1.00	0.01
Affinity+EM	0.01	0.38	0.00

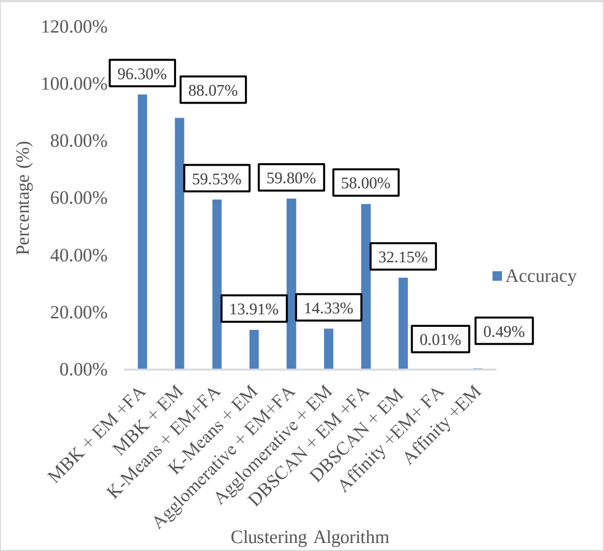


Fig. 2. Comparison of Accuracy Performance.

Fig. 2 shows the accuracy score of all algorithms. Overall, most algorithms have shown an increase in accuracy when a hybrid with FA. Only Affinity+EM is excluded. The proposed MBK+EM+FA has increased by 2.9%. It is revealed that the proposed algorithm is more accurate and capable of converging compared to other algorithms. It is observed that the f-measure, Precision, recall, and accuracy of the proposed clustering algorithm are the highest values. These values indicate that the performance of the proposed clustering algorithm was satisfactory based on high-performance parameters achieved.

Table III shows the comparative performance of the proposed clustering algorithm compared with other hybrid clustering algorithms based on external validation in terms of homogeneity, FMI, completeness, and V-Measure. Interestingly, all five clustering algorithms mostly performed better with the hybrid FA. All measurements have demonstrated much increment. It shows that clustering validation on homogeneity agreement, the perfectness of labeling, and independent ability in clustering and clustering agreement is acceptable. However, the most efficient clustering algorithm is the proposed MBK+EM +FA. As seen in the table, it is evident that the increase of about 0.3 compared with MBK+EM for all external validation measurements. Most of the values are between 0.78 and 0.94. The highest FMI value is 0.9307. The obtained FMI indicates a good clustering agreement offered by the proposed MB+EM+FA due to the value being almost zero.

The time consumption of the algorithm is defined as the time required to assess all the data to generate clusters. The elapsed time of the algorithms is calculated and expressed in seconds, as shown in Table IV. Table IV indicates that the proposed MBK+EM+FA that uses entropy measure and FA consumes less time than K-Means+EM+FA, Agglomerative+ EM+FA, DBSCAN+EM+FA, and Affinity+EM+FA. It reveals that the hybrid proposed algorithm is much more effective than others. However, the use of MBK+EM+FA is a bit higher in computational time compared to MBK+EM. However, as mentioned in the previous section, it offers a good accuracy performance. A slight increase in execution time does not cause an issue in the practical use of the algorithm because a better performance is obtained. Overall, the computational time is less than a minute. Thus, it can be said to be sufficient time for the execution of cluster data since the proposed algorithm involves a searching mechanism compared to the others [29][30].

TABLE III. CLUSTERING PERFORMANCE COMPARISON FOR HOMO, FMI, COMP AND VM

Clustering Algorithms	**HOMO**	**FMI**	**COMP**	**VM**
MBK+EM+FA	0.7914	0.9307	0.7804	0.7858
MBK+EM	0.4777	0.804	0.4576	0.4673
K-Means+ EM+FA	0.0006	0.7192	0.0639	0.0013
K-Means+EM	0.4479	0.7742	0.4218	0.4343
Agglomerative+EM+FA	0.0006	0.7192	0.0639	0.0013
Agglomerative+EM	0.4496	0.7679	0.4214	0.4349
DBSCAN+EM+FA	0.0090	0.6938	0.0324	0.0141
DBSCAN+EM	0.0005	0.7033	0.0392	0.0002
Affinity+EM+FA	0.8493	0.1285	0.1232	0.2152
Affinity+EM	0.729	0.1178	0.101	0.1699

TABLE IV. EXECUTION TIME (PER SECONDS)

Clustering Algorithms	**Time (s)**
MBK+EM+FA	0.4589
MBK+EM	0.3550
K-Means+ EM+FA	0.5839
K-Means+EM	0.7850
Agglomerative+EM+FA	14.1792
Agglomerative+EM	13.9627
DBSCAN+EM +FA	87.7652
DBSCAN+EM	8.3744
Affinity+EM+FA	23.03922
Affinity+EM	11.7965

VII. DISCUSSION

The capability and effectiveness of a clustering approach to reduce the clustering error and improve the accuracy are the most crucial qualities in clustering. There is no inherent distance between the feature of categorical data analysis remains challenging. This research aims to evaluate the clustering efficiency of heterogeneous categorical data. Entropy is a measure of information. The entropy distance generates within the clustering algorithm approach allows us to measure the distance of features systematically quantified. The entropy aided the clustering algorithms in selecting the center of the centroid. Due to the nature of the entropy, it can investigate the compatibility of data to produce weighted values that can represent each class in the dataset. The initial weight value of entropy can study the data well and produces a higher accuracy.

Introducing the FA approach in the proposed clustering algorithms establishes a new search strategy for finding a globally optimal solution. The aim is to find the nearest features based on the entropy distance measure. In this case, the total number of categorical features is 111 features. Overall, achieving a clustering algorithm embedded with FA provides a good performance. It supports the previous work FA can contribute to an efficient solution due to its flexibility, simplicity, and fast convergence [31]. FA also offers a global search strategy [17] that can explore more search spaces for all features in finding the nearest neighbor. The generated value for the entropy measure seems to improve when using the FA. It could reflect in the clustering similarity measure based on a data point in both the intra-distance and inter-distance among the features. As illustrated in the above results, the best performance is by the proposed MBK+EM+FA. However, the high quality of the clustering results in more than constitutes for such restrictions of elapsed time. In addition, more work can be done such as embedding the *k*-interpolation model based on Kriging Method [25] and other computational optimization methods.

VIII. CONCLUSION

This research compares clustering algorithms that utilize the fundamental behavior of the firefly algorithm in order to improve clustering problem-solving. The proposed firefly algorithm can locate the cluster centers by comparing several clustering techniques to refine the centers as input. The experimental results have proved the effectiveness, capability, and efficiency of the FA and MBK with EM in improving the clustering performance in heterogeneous categorical data of public perception of timber utilization as compared to MBK+EM, K-Means+EM+FA, K-Means+EM, Agglome rative+EM+FA, Agglomerative+EM, DBSCAN+EM +FA, DBSCAN+EM, Affinity+EM+ FA, and Affinity+EM. In improving the solution, other objective functions considering inter and intra-cluster measurements could be used in future research to improve the proposed models. Another aspect is the evaluation of heterogeneous categorical data from a different domain.

ACKNOWLEDGMENT

The authors would like to acknowledge the Ministry of Higher Education (Fundamental Research Grant Scheme (FRGS) Grant: 600-IRMI/FRGS 5/3 (213/2019)), Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA and Malaysian Timber Industry Board, Malaysia for the financial support provided to this research project.

REFERENCES

[1] Sapegin and C. Meinel, “K-metamodes: Frequency-and ensemble-based distributed k-modes clustering for security analytics,” Proc. - 19th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2020, pp. 344–351, 2020, doi: 10.1109/ICMLA51294.2020.00062.

[2] S. V Ambadkar and S. P. Akarte, “Clustering Categorical Data for Internet Security Applications,” Int. J. Sci. Tech. Adv., vol. 2, no. 1, pp. 115–118, 2016.

[3] N. Ifada, M. E. Ariyanto, M. K. Sophan, and M. Nikmat, “A Comparative Study of Centroid and Medoid based Categorical Data Clustering Methods for Solving Cold-start Recommendation Problem,” CENIM 2020 - Proceeding Int. Conf. Comput. Eng. Network, Intell. Multimed. 2020, pp. 418–422, 2020, doi: 10.1109/CENIM51130. 2020.9297960.

[4] Y. V. Via et al., “Optimization of Production Profits Using The Firefly Algorithm,” pp. 291–296, 2020.

[5] A. Hassan and T. M. Tawfeeg, “Greedy Firefly Algorithm for Optimizing Job Scheduling in IoT Grid Computing,” pp. 1–18, 2022.

[6] B. Filipowicz, “Firefly algorithm in optimization of queueing systems,” vol. 60, no. 2, pp. 363–368, 2012, doi: 10.2478/v10175-012-0049-y.

[7] V. S. Rajput, “A New Approach of Firefly Algorithm for Optimizing Reviews of Opinion Mining,” pp. 18–23, 2016.

[8] I. T. R. Yanto, Y. Saadi, D. Hartama, D. P. Ismi, and A. Pranolo, “A framework of fuzzy partition based on Artificial Bee Colony for categorical data clustering,” 2nd Int. Conf. Sci. Inf. Technol., pp. 260–263, 2017, doi: 10.1109/ICSITech.2016.7852644.

[9] K. Lakshmi, N. Karthikeyani Visalakshi, S. Shanthi, and S. Parvathavarthini, “CLUSTERING CATEGORICAL DATA USING k-MODES BASED ON CUCKOO SEARCH OPTIMIZATION ALGORITHM,” ICTACT J. Soft Comput., vol. 8, no. 1, pp. 1561–1566, 2017, doi: 10.21917/ijsc.2017.0218.

[10] H. Wu and S. O. Leung, “Can Likert Scales be Treated as Interval Scales?—A Simulation Study,” J. Soc. Serv. Res., vol. 43, no. 4, pp. 527–532, Aug. 2017, doi: 10.1080/01488376.2017.1329775.

[11] F. Cao et al., “An Algorithm for Clustering Categorical Data with Set-Valued Features,” IEEE Trans. Neural Networks Learn. Syst., vol. 29, no. 10, pp. 4593–4606, 2018, doi: 10.1109/TNNLS.2017.2770167.

[12] T. Hien, T. Nguyen, D. Tai, D. Songsak, and S. Van Nam, “A method for k-means-like clustering of categorical data,” J. Ambient Intell. Humaniz. Comput., no. Berkhin 2002, 2019, doi: 10.1007/s12652-019-01445-5.

[13] Y. Zhang, Y. M. Cheung, and K. C. Tan, “A Unified Entropy-Based Distance Metric for Ordinal-and-Nominal-Attribute Data Clustering,” IEEE Trans. Neural Networks Learn. Syst., vol. 31, no. 1, pp. 39–52, Jan. 2020, doi: 10.1109/TNNLS.2019.2899381.

[14] H. Sun, R. Chen, S. Jin, and Y. Qin, “A hierarchical clustering for categorical data based on holo-entropy,” Proc. - 2015 12th Web Inf. Syst. Appl. Conf. WISA 2015, pp. 269–274, 2016, doi: 10.1109/WISA.2015.18.

[15] Y. Xiao, C. Huang, J. Huang, I. Kaku, and Y. Xu, “Optimal mathematical programming and variable neighborhood search for k-modes categorical data clustering,” Pattern Recognit., vol. 90, no. Huang 1997, pp. 183–195, 2019, doi: 10.1016/j.patcog.2019.01.042.

[16] E. J. Rivera Rios, M. A. Medina-Pérez, M. S. Lazo-Cortés, and R. Monroy, “Learning-based dissimilarity for clustering categorical data,” Appl. Sci., vol. 11, no. 8, pp. 1–17, 2021, doi: 10.3390/app11083509.

[17] A. E. S. Ezugwu, M. B. Agbaje, N. Aljojo, R. Els, H. Chiroma, and M. A. Elaziz, “A Comparative Performance Study of Hybrid Firefly Algorithms for Automatic Data Clustering,” IEEE Access, vol. 8, pp. 121089–121118, 2020, doi: 10.1109/ACCESS.2020.3006173.

[18] H. Xie et al., “Improving K-means Clustering with Enhanced Firefly Algorithms,” no. September, 2019, doi: 10.1016/j.asoc.2019.105763.

[19] A. Jaradat, B. Matalkeh, and W. Diabat, “Solving Traveling Salesman Problem using Firefly algorithm and K-means Clustering,” 2019 IEEE Jordan Int. Jt. Conf. Electr. Eng. Inf. Technol. JEEIT 2019 - Proc., pp. 586–589, 2019, doi: 10.1109/JEEIT.2019.8717463.

[20] M. Takeuchi, T. Ott, H. Matsushita, Y. Uwate, and Y. Nishio, “K-Means Clustering with Modifying Firefly Algorithm,” Int. Symp. Nonlinear Theory Its Appl., no. 1, pp. 576–579, 2017.

[21] A. Saha and S. Das, “Categorical fuzzy k-modes clustering with automated feature weight learning,” Neurocomputing, vol. 166, pp. 422–435, 2015, doi: 10.1016/j.neucom.2015.03.037.

[22] X. S. Yang and X. He, “Firefly algorithm: recent advances and applications,” Int. J. Swarm Intell., vol. 1, no. 1, p. 36, 2013, doi: 10.1504/ijsi.2013.055801.

[23] Meghana M Chavan, Asawari Patil, Lata Dalvi, and Ajinkya Patil, “Mini Batch K-Means Clustering On Large Dataset,” Int. J. Sci. Eng. Technol. Res., vol. 04, no. 07, pp. 1356–1358, 2015.

[24] A. Feizollah, N. B. Anuar, R. Salleh, and F. Amalina, “Comparative study of k-means and mini batch k-means clustering algorithms in android malware detection using network traffic analysis,” Proc. - 2014 Int. Symp. Biometrics Secur. Technol. ISBAST 2014, pp. 193–197, 2015, doi: 10.1109/ISBAST.2014.7013120.

[25] S. Bonnini, “Testing for heterogeneity with categorical data: Permutation solution vs. bootstrap method,” Commun. Stat. - Theory Methods, vol. 43, no. 4, pp. 906–917, 2014, doi: 10.1080/03610926.2013.799376.

[26] X. Yang, Nature-Inspired Metaheuristic Algorithms. 2010.

[27] E. M. Mashhour, E. M. F. El Houby, K. T. Wassif, and A. I. Salah, “Feature selection approach based on firefly algorithm and chi-square,” Int. J. Electr. Comput. Eng., vol. 8, no. 4, pp. 2338–2350, 2018, doi: 10.11591/ijece.v8i4.pp2338-2350.

[28] A. Karim, S. Azam, B. Shanmugam, and K. Kannoorpatti, “An Unsupervised Approach for Content-Based Clustering of Emails into Spam and Ham through Multiangular Feature Formulation,” IEEE Access, vol. 9, pp. 135186–135209, 2021, doi: 10.1109/ACCESS.2021.3116128.

[29] A. Rauf, Sheeba, S. Mahfooz, S. Khusro, and H. Javed, “Enhanced K-mean clustering algorithm to reduce number of iterations and time complexity,” Middle East J. Sci. Res., vol. 12, no. 7, pp. 959–963, 2012, doi: 10.5829/idosi.mejsr.2012.12.7.1845.

[30] M. P. Behera, A. Sarangi, and D. Mishra, “K-medoids crazy firefly algorithm for unsupervised data clustering,” 1st Odisha Int. Conf. Electr. Power Eng. Commun. Comput. Technol. ODICON 2021, 2021, doi: 10.1109/ODICON50556.2021.9428980.

[31] N. Dey, Springer Tracts in Nature-Inspired Computing Applications of Firefly Algorithm and its Variants Case Studies and New Developments. Kolkata, West Bengal, India, 2020.

Improved Spatial Invariance for Vehicle Platoon Application using New Pooling Method in Convolution Neural Network

M S Sunitha Patel[1], Srinath S[2]

Dept. of Computer Science & Engineering, ATME College of Engineering, Mysuru, Karnataka, India[1]
Dept. of Computer Science & Engineering, SJCE College of Engineering, Mysuru, Karnataka, India[2]

***Abstract*—The imbalanced dataset is a prominent concern for automotive deep learning researchers. The proposed work provides a new mixed pooling strategy with enhanced performance for imbalanced vehicle dataset based on Convolution Neural Network (CNN). Pooling is crucial for improving spatial invariance, processing time, and overfitting in CNN architecture. Max and average pooling are often utilized in contemporary research articles. Both techniques of pooling have their own advantages and disadvantages. In this study, the advantages of both pooling algorithms are evaluated for the classification of three vehicles: car, bus, and truck for imbalanced datasets. For each epoch, the performance of max pooling, average pooling, and the new mixed pooling method was assessed using ROC, F1-score, and error rate. Comparing the performance of the max-pooling method to that of the average pooling method, it has been found that the max-pooling method is superior. The performance of the proposed mixed pooling approach is superior to that of the maximum pooling and average pooling methods. In terms of Receiver Operating Characteristics (ROC), the proposed mixed pooling technique is approximately 2 per cent better than the maximum pooling method and 8 per cent better than the mixed pooling method. Using a new pooling technique, the classification performance with an imbalanced dataset is improved, and also a novel mixed pooling method is proposed for the classification of vehicles.**

Keywords*—*Average pooling; convolution neural network; imbalance dataset; max pooling; mixed pooling

I. Introduction

Automotive companies are conducting cutting-edge research on vehicle platoon management to improve 'vehicle-to-vehicle' communication for enhanced performance. Vehicle platoon management increases customer and societal benefits by attaining greater fuel efficiency, less pollution, less road congestion, and fewer road accidents [1] to [2]. To improve performance, a key feature of vehicle platoon management is categorizing vehicles based on their size and grouping them suitably to achieve reduced aerodynamic drag when driving in the longitudinal direction, as illustrated in Fig. 1. To achieve the requirement, vehicles having Advanced Driving Assistance System (ADAS) features such as lane-keeping, automated cruise control, pedestrian safety, platoon management, and others must have at least two cameras installed. Vehicle platoon management is also categorized as ADAS [3], which falls within the L2 and L3 levels of vehicle automation. As vehicle automation level increases from L1 to L5 then vehicle intelligence should also increase to improve safety. Considering the features given by automotive manufacturers for modern vehicles, it is clear that camera utilization is expanding day by day, necessitating further study into image processing algorithms. Strict safety requirement calls for the need for more cameras in vehicles along with other sensors. As the number of cameras on a vehicle increases, image processing output will become more vital for controlling the vehicle according to safety requirements.

CNN has been widely used in vehicle image processing applications like pedestrian safety [4], vehicle classification [5,6,7], and many more applications. The focus of current research is on categorizing vehicles by type, which is the initial stage in vehicle platoon management before combining them for greater efficiency. To begin with car, bus, and truck are three classes of vehicles that are regularly seen on highways, and these three have been taken into consideration for experimental work using CNN. To train the CNN algorithms, high-quality and sufficient volumes of images are required. Unlike other applications where image datasets are publicly available, such as medical imaging, sign language identification, pattern recognition, and so on, the availability of vehicle image datasets is limited. Open repository datasets may not contain all of the images needed to do experimental work on the specified topic. In another instance, researchers have created datasets for vehicle rear parts in [8], and a few vehicle datasets are published in open source PKU-VD [9], VeRi-776 [10], VehicleID [11], as well as different vehicle datasets are available for smart city study [12]. But, to get dataset access from open source, researchers need to get approval from the owners. However, researchers must obtain permission from the owners of open-source datasets to access them. Still, sometimes delays or no response can be expected when seeking approval from the owners. There are few open access datasets for automotive applications [14,15,16]. While developing required image dataset, all this leads to an imbalanced image dataset collection, which is a prevalent problem in the automotive arena. In multiclass analysis, imbalance datasets indicate that the dataset's distribution includes unequal amounts of images for each class. Uneven quantities of images will impact the accuracy of the learning algorithms. In comparison to the amount of car images, the availability of images in open source for bus and truck images are minimal. As a result, various image augmentation techniques such as flipping, rotation, cropping, and so on are used to increase the quantity of bus and truck images. Therefore, to increase the classification performance even with an imbalanced dataset a mixed pooling approach has been proposed in this research. In the further sections deal with the works carried out related to vehicle classification using CNN and the motivation behind the mixed pooling implementation to improve the performance of the CNN learning algorithm for the imbalance vehicle datasets has been explained.

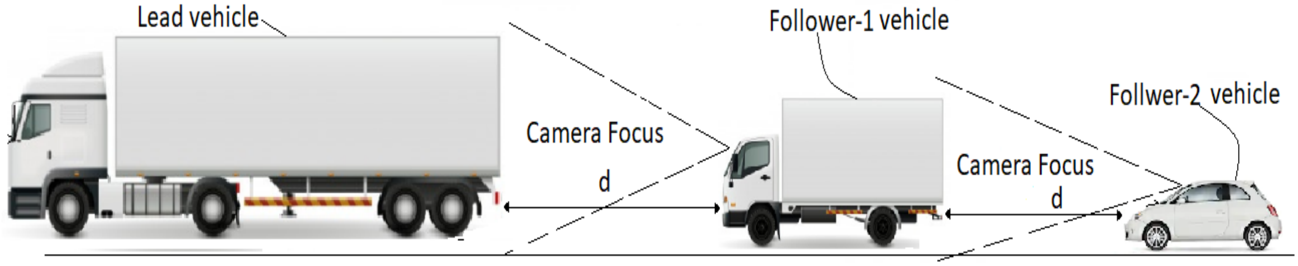


Fig. 1. Example for Three Vehicles in Platoon Management, Where the First Vehicle is called as Lead Vehicle and subsequent Two Vehicles are called as Follower-1 and Follower-2. Vehicles Separated from each with Distance - d.

II. Related Work

Deep Learning methods for image processing currently dominate computer vision, particularly for image recognition related applications. Several works utilising Deep Learning for vehicle detection are documented in the literature.

Using the CNN deep learning technique, [23] proposed a daytime vehicle detection system. Using appearance-based feature extraction techniques, the experiment yielded improved results. Fast R–CNN-based vehicle classification algorithm for real-time traffic surveillance was developed by Wang et al. [24]. A dataset of 60,000 images depicting traffic junction was compiled and divided into training and tested data, on which the suggested approach achieved an accuracy of 80.051%. Chauhan et al. [25] have developed a CNN-based framework for the classification and counting of vehicles on highways. On the collected dataset of 5,562 CCTV camera videos of highway traffic, the suggested framework achieved 75 percent MAP. Jo et al. [13] have suggested a GoogLeNet framework for vehicle categorization based on transfer learning. The authors demonstrated that the provided classifier achieved a 0.983% accuracy rate on the ILSVRC-2012 data set. Chia-Chi Tsai et al., 2018 [17], developed an improved Convolutional Neural Network architecture based on deep learning methods for intelligent transportation applications. Improved Spatio-Temporal Sample Consensus is the name of a method developed by Yu Wang et al. [26] for detecting and classifying moving vehicles. First, the moving vehicles are detected using the Spatio Temporal Sample Consensus technique, based on the brightness variation and shadow of the vehicles. In addition, using feature fusion algorithms, the objects are categorised based on area, face, licence plate, and vehicle symmetry characteristics. In the study proposed by Kaiming He [18], the deep networks are equipped with the pooling approach called spatial pyramid pooling. This eliminates the requirement of fixed size image being given as input to the network. The novel network structure, termed SPP-net, may provide a fixed-length representation independent of image size/scale. Pyramid pooling is also robust to object deformations. With these features, SPP-net improves all CNN-based image classification algorithms.

Deep feature-based techniques can effectively improve the accuracy of vehicle classification, but they require an enormous quantity of data to attain considerable accuracy. Hence, in this current work an attempt has been made to achieve good accuracy with moderate dataset size.

III. Motivation

CNN's components include a convolution layer, a pooling layer, and a flattening layer that performs feature extraction and uses the output of the flattening vector for classification. The convolution layer performs a linear operation that extracts a feature map from an image. For the given image, the ith convolution layer generates ith output feature map and it can be represented as mentioned in (1) where ‘w_i’ is kernel or filter and ‘x’ is the input image with 2D convolution operator and y is an output.

$$y_i = f(w_i * x) \tag{1}$$

The kernel used in convolution will travel across the image left to right and top to bottom, the amount of movement for the given input image depends on the input image size and kernel size. The activation function in the convolution layer produces a feature map, which is fed into the pooling layer. Pooling is a nonlinear function that generates output by summing the net of certain pixel positions. The pooling layer has numerous advantages in CNN architecture, including improved processing time, noise invariance, and overfitting. In the following sections below some of the most commonly utilized pooling strategies are discussed.

A. Max Pooling

Max Pooling selects the maximum value in the feature map for the selected filter dimension. Here computation has made faster [19, 20] because of the elimination of non- max values and the output of the max pool will reduce from X to Y as mentioned in Fig. 2. By selecting the max value in the feature map, max-pooling picks the brightest pixel from the given image. Max pooling can be represented mathematically as given in (2). Where for given ith feature map x is the input element at (p,q) within pooling region R_{jk} which represents local neighbourhood position (j,k)with output y.

$$y_{ijk} = \max_{(p,q)\in R_{jk}} (x_{ipq}) \tag{2}$$

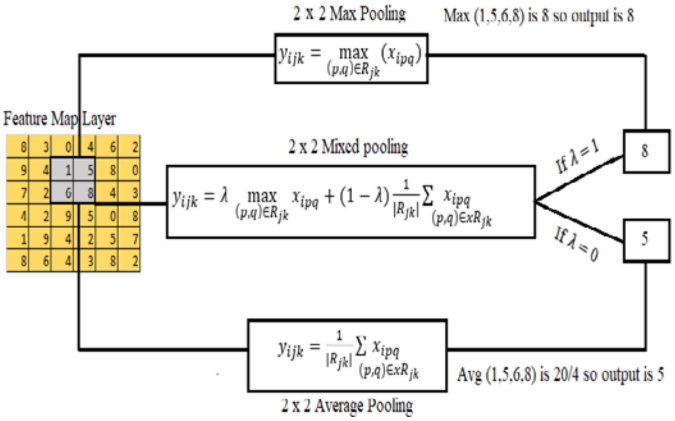

2 x 2 Max Pooling
Max (1,5,6,8) is 8 so output is 8
$y_{ijk} = \max_{(p,q)\in R_{jk}} (x_{ipq})$
Feature Map Layer
8 3 0 4 6 2
9 4 1 5 8 0
7 2 6 8 4 3
4 2 9 5 0 8
1 9 4 2 5 7
8 6 4 3 8 2
2 x 2 Mixed pooling
$y_{ijk} = \lambda \max_{(p,q)\in R_{jk}} x_{ipq} + (1-\lambda)\frac{1}{|R_{jk}|}\sum_{(p,q)\in xR_{jk}} x_{ipq}$
If λ=1
8
If λ=0
5
$y_{ijk} = \frac{1}{|R_{jk}|}\sum_{(p,q)\in xR_{jk}} x_{ipq}$
Avg (1,5,6,8) is 20/4 so output is 5
2 x 2 Average Pooling

Fig. 2. Toy Example to Illustrate 2x2 Max, Average and Mixed Pooling Principle.

B. *Average Pooling*

Average pooling selects the average value in the feature map for the selected filter dimension. The computation is faster [21] since it selects the average value of each region and the output value is lowered from X to Y, as illustrated in Fig. 2. By selecting the average value in the feature map, average pooling picks the average pixel from the given image. Average pooling is represented mathematically as given in (3). Where for given ith feature map x is the input element at (p,q) within pooling region R_{jk} which represents local neighbourhood position (j,k)with output y.

$$y_{ijk} = \frac{1}{|R_{jk}|}\sum\nolimits_{(p,q)\in xR_{jk}} x_{ipq} \quad (3)$$

C. *Mixed Pooling*

Max pooling performance will degrade when the feature map has a low pixel value and average pooling performance will degrade when the feature map has high pixel values. To overcome this problem, mixed pooling approach has been implemented [22]. As illustrated in Fig. 2, the mixed pooling approach combines both max and average pooling by picking any one method during execution. Mixed pooling can be represented as in (4). Where λ can be 1 or 0 and is chosen at random; pooling strategy is average if λ is 0 and pooling strategy is to max if λ is 1.

$$y_{ijk} = \lambda \ \max\nolimits_{(p,q)\in R_{jk}} x_{ipq} + (1-\lambda)\frac{1}{|R_{jk}|}\sum\nolimits_{(p,q)\in xR_{jk}} x_{ipq} \quad (4)$$

D. *Proposed Novel Method*

In the current work, novel pooling method has been proposed to get the benefits of mixed and max pooling. Because mixed pooling has the advantage of taking into account both the max and average value of the pixels, it has been implemented after the first layer of convolution, and max pooling has been implemented in the second and third layer pooling, as shown in Fig. 3 for a three-layer CNN. The proposed method has the advantage that once the first convolution layer captures the maximum needed feature using mixed pooling; the subsequent layer using max pooling will reinforce the features.

IV. Methodology

The present work begins with gathering bus, car, and truck images and creating the requirement dataset for the vehicle platoon environment. With the developed dataset training, testing and validation has been performed for three different pooling scenarios as presented in Fig. 3. Performance evaluation for all three pooling scenarios has been visualized using ROC Curve.

A. *Dataset Creation*

To undertake experimental work in deep learning, the first step is to develop a suitable image dataset to train the deep learning algorithm. The car images were acquired from an open-source repository [14], the bus images were acquired from an open-source repository [15], and truck images were acquired from the open-source repository [16]. There were approximately 3000 images available for truck and bus, and nearly 8000 images were acquired for the car vehicle class. The image data size for bus and truck has been increased from 3000 images to 5000 images using image augmentation techniques. Different image augmentation techniques like padding, flip, rotate, tilt, blur, crop, adding noise, etc. have been incorporated based on the available images. Therefore, the total number of images used to conduct the experiment is 5000 for the bus and truck classes and 8000 for the vehicle class.

B. *Training, Testing and Validation Dataset*

As demonstrated in Fig. 4 by utilizing open source and different augmentation techniques final dataset has been created. The training and testing ratio are set at 80% and 20%, respectively. To optimize the CNN algorithm's learning performance, 20% validation datasets were used.

C. *CNN Architecture*

As mentioned earlier, a three-layer Convolution Neural Network has been implemented as shown in Fig. 5. The experiment has been carried out using Python 3.7.9, with IDE PyCharm and OpenCV Library. The CNN architecture has taken care of feature extraction, including filtering, image segmentation, and image enhancement for images that are 64 by 64 in size.

CNN model consists of convolution layer, pooling layer, fully connected layer, and output layer. The first convolution layer consists of 32 filters, followed by 64 filters in the second and third layers. Padding has been considered to maintain input and output image size. Three alternative pooling strategies have been studied in terms of pooling techniques. In Scenario-1, all three pooling configurations are max pooling. In Scenario-2, all three pooling configurations are mixed pooling and in Scenario-3 the first pooling configuration is mixed pooling and the second two pooling configurations are max pooling. The output of pooling with window size 2 and stride size 2 is fed to the size 2 of the fully connected layer. The fully connected layer is the deep layer which is used for classification. Finally, the fully connected layer passes its input to softmax with 2 layers for the final classified image output. Each scenario's output is executed in three distinct steps, and the outcomes are summarized and discussed in the following section.

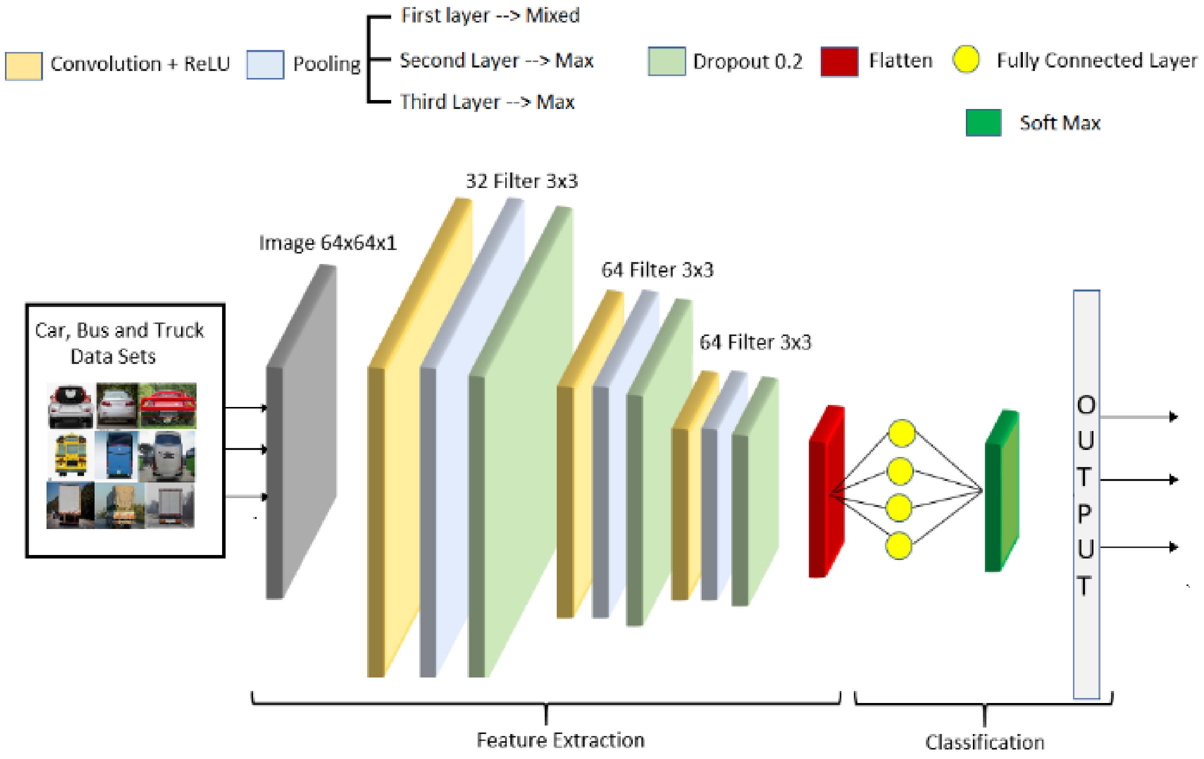


Fig. 3. Proposed Architecture Where the First Pooling Layer will be mixed and the subsequent Pooling Layer will be Max.

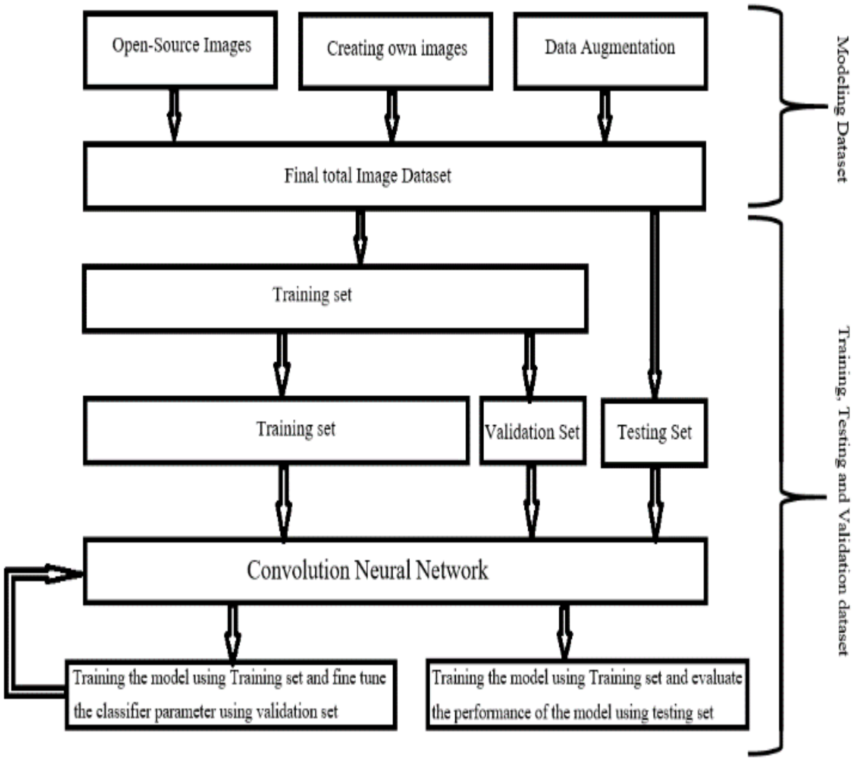
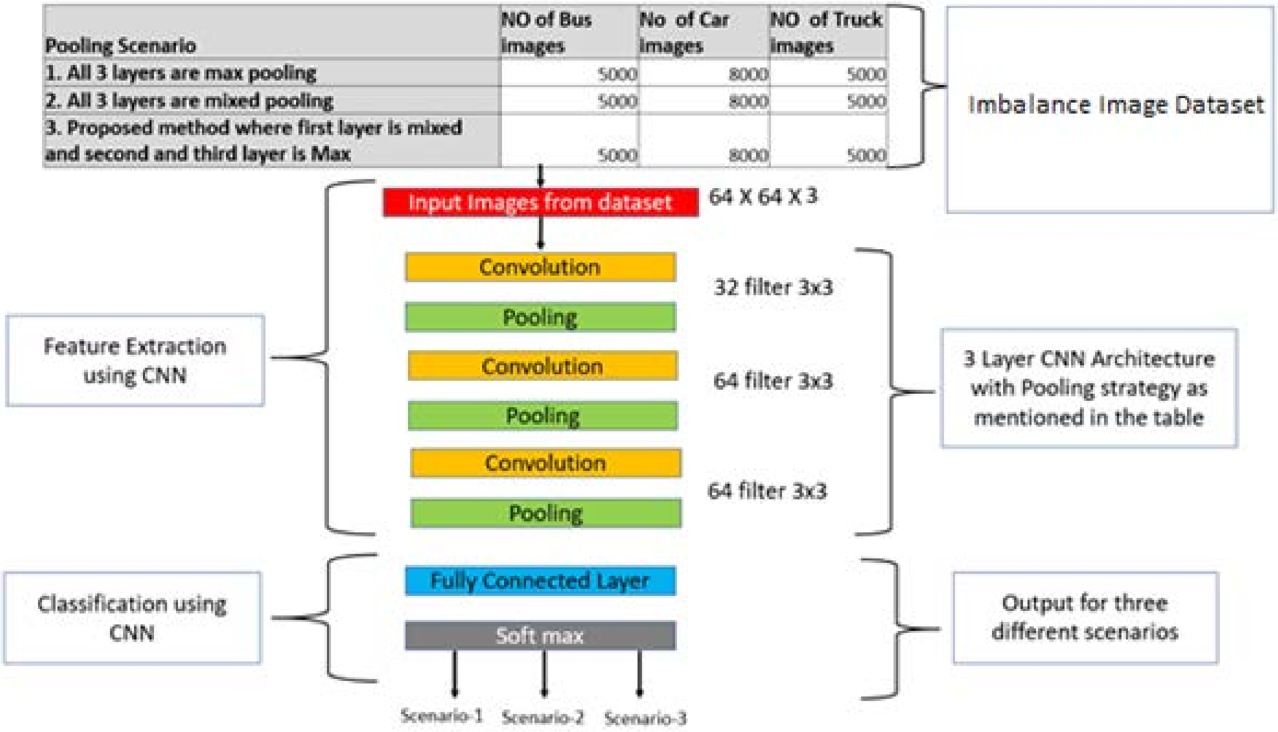

Open-Source Images
Creating own images
Data Augmentation
Final total Image Dataset
Modeling Dataset
Training set
Training set
Validation Set
Testing Set
Convolution Neural Network
Training the model using Training set and fine tune the classifier parameter using validation set
Training the model using Training set and evaluate the performance of the model using testing set
Training, Testing and Validation dataset

Fig. 4. Training Dataset 80% and Testing Dataset 20 %, for Fine-tuning the Learning Algorithm 20% of the Validation Dataset has been used.



Pooling Scenario	NO of Bus images	No of Car images	NO of Truck images
1. All 3 layers are max pooling	5000	8000	5000
2. All 3 layers are mixed pooling	5000	8000	5000
3. Proposed method where first layer is mixed and second and third layer is Max	5000	8000	5000

Fig. 5. Implementation Methodology.

D. Performance Evaluation

Standard formulae of accuracy, precision, recall, and F1 score, as indicated (5) through (8), were used to evaluate the performance of three distinct pooling procedures. Here, True Positive (TP) is a correctly predicted class, False Positive (FP) is a label that does not belong to class but is predicted as positive, True Negative (TN) is the correctly predicted for class that does not belong to the class, False Negative (FN) is wrongly predicted for class that does not belong to the class.

$$\text{Accuracy} = \frac{(\text{TP + TN})}{(\text{TP + TN+ FP+ FN})} \quad (5)$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+ FP })} \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP + FN })} \quad (7)$$

$$\text{F1 Score} = \frac{2 \text{ x TP}}{((2 \text{ x TP}) + \text{FP+ FN })} \quad (8)$$

V. Results and Discussion

As depicted in Fig. 5. imbalance image datasets were fed into the CNN model and run individually for each of the three cases and the results obtained are as shown in Fig. 6. In Fig. 6, label 0 denotes the bus class, label 1denotes the car class, and label 2 denoted the truck class. ROC graphs are also being used to visualize the results of Scenario-1, Scenario-2, and Scenario-3 as illustrated in Fig. 7, Fig. 8, and Fig. 9, respectively.

A. Scenario-1(All Three Max Pooling)

Fig. 7 illustrates the accuracy attained for all three classes using the ROC curve for Scenario-1. The acquired findings show that the achieved average accuracy is 98 % with 100 % accuracy for vehicle classification and 95 % and 97 %accuracy for bus and truck classification. After execution, false-positives for bus and truck are 150 and 250 images, respectively, and false-negatives are 250 and 150 images for the 20% overall dataset during testing.

B. Scenario -2 (All Three Mixed Pooling)

The ROC curve for Scenario-2 is shown in Fig. 8. The average accuracy achieved is 98%, with 100% classification for car class and 94 % and 97 % for bus and truck class, respectively. The false-positive rate for bus and truck is 150 and 300 images, respectively, whereas the false-negative rate for the 20% overall dataset during testing is 300 and 150 images.

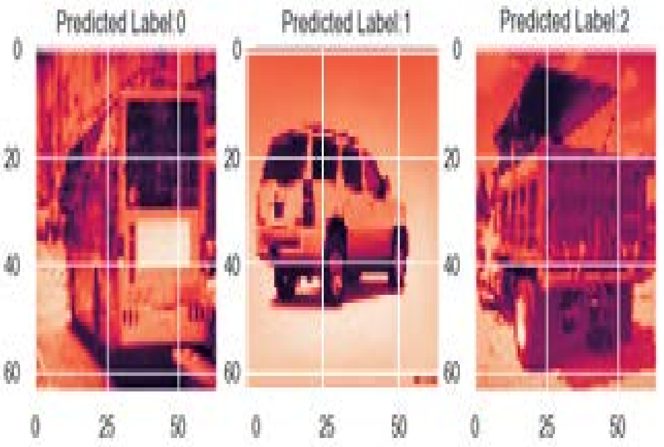


Fig. 6. Predicted Output (Label Convention: Label 0 -Bus, Label 1- Car, Label 2-truck).

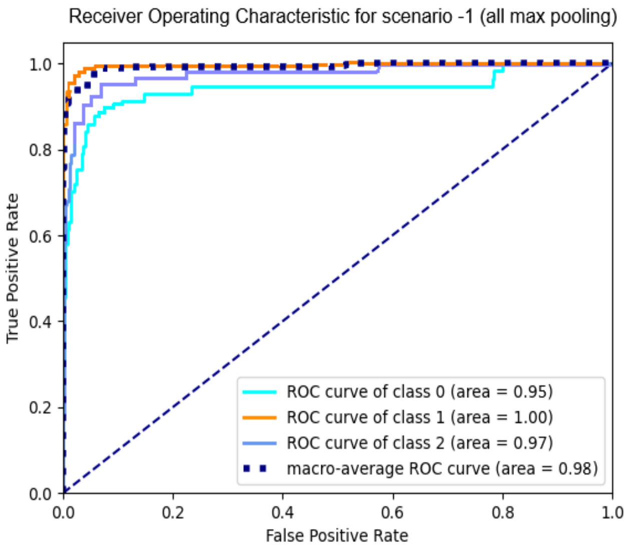


Fig. 7. ROC for Scenario-1, Where Class 0 is for Bus, Class 1 is for Car and Class 2 is for Truck.

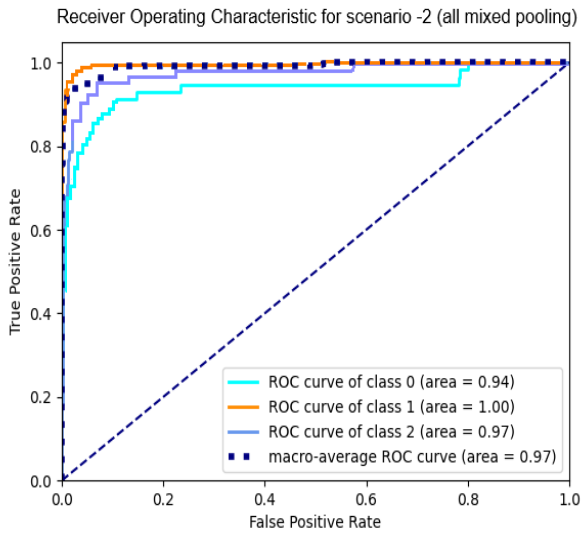


Fig. 8. ROC for Scenario-2, where Class 0 is for Bus, Class 1 is for Car and Class 2 is for Truck.

C. Scenario-3 (New Mixed Pooling)

Fig. 9 depicts the ROC curve for Scenario-3. While 100% of cars are correctly classified, 94% and 97% of buses and trucks are correctly classified. The number of false-positive images for the bus and truck is 150 and 50 respectively and the number of false-negative images is 50 and 150, respectively for the 20% overall dataset during testing.

D. Epoch Error Rate

From Fig. 10, it can be observed that the learning cycle has been completed nearly at 50 epochs for all three scenarios. However, in the proposed methodology, the pooling error rate is decreased to 0.7 percent, whereas the error rate in max-pooling is 1% and 1.4 per cent in mixed pooling.

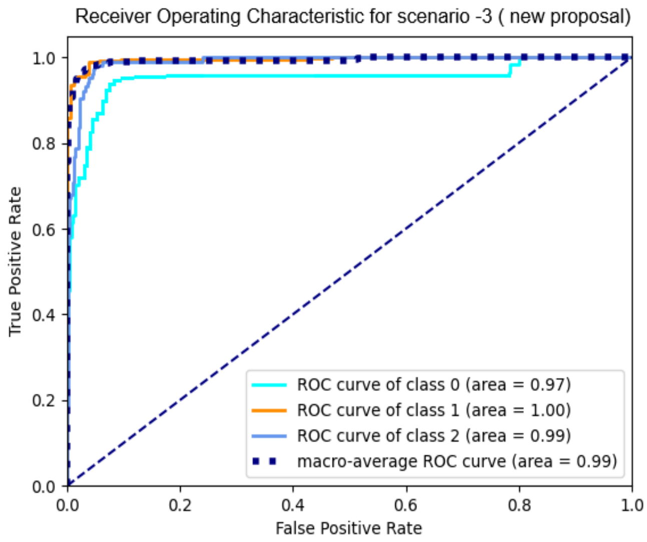


Fig. 9. ROC for Scenario-3, Where Class 0 is for Bus, Class 1 is for Car and Class 2 is for Truck.

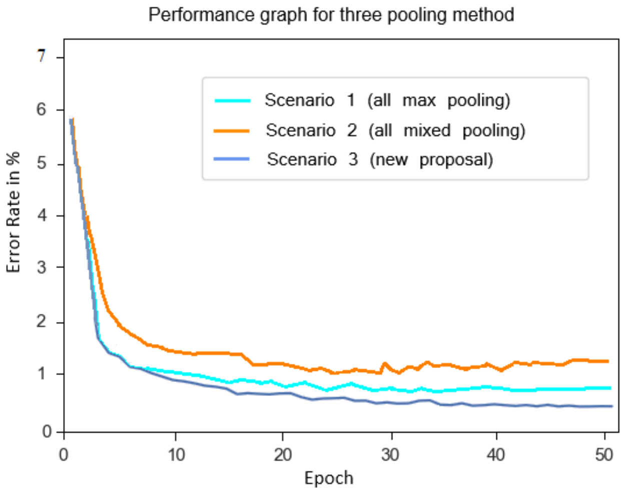


Fig. 10. Performance at different Epochs of for Three different Pooling Scenarios.

E. Overall Performance Evaluation

As the literature review reveals no evidence of deep learning based vehicle grouping to aid in vehicle platoon management, an attempt has been made where the proposed methodology enhances the performance of vehicle classification using an imbalanced image dataset. For the three different situations that were taken into consideration for the experimentation, Table I list the performance metric values for accuracy, precision, recall, and f1-score for multiclass vehicle classification. From the experimental outcomes, it can be inferred that the suggested strategy outperformed other methods while using its own dataset for the platoon management system.

To demonstrate the improved performance of the proposed method, the results obtained from the proposed method is compared with the other methods as shown in Table II. The proposed method also shows a consistent classification performance for all the vehicle classes.

TABLE I. SUMMARY OF THE PERFORMANCE EVALUATION FOR ALL THREE SCENARIOS

Class	**Performance Metric**	**Scenari-1 output in %**	**Scenari-2 output in %**	**Scenari-3 output in %**
BUS	Accuracy	95	94	97
	Precision	97	97	99
	Recall	95	95	97
	F1 Score	96	95	98
Car	Accuracy	100	100	100
	Precision	100	100	100
	Recall	100	100	100
	F1 Score	100	100	100
Truck	Accuracy	97	97	99
	Precision	95	94	97
	Recall	97	97	97
	F1 Score	96	96	98
Bus, Car & Truck	Average Accuracy in % for all 3 class	98	97	99
Bus, Car & Truck	Average error rate in % at epoch 50	1.4	1	0.7

TABLE II. PERFORMANCE COMPARISON WITH PROPOSED METHOD TO OTHER METHODS

Method	**Vehicle class considere d**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
F-RCNN (8)	Bike , Bus and Truck	94.4	88.6	93.5	88
YOLO (7)	Bus, Car, and Truck	99	Not applicable	Not applicable	Not applicable
CNN-Super learner (6)	Bus, Car, Truck, Pedestrian, and Bike	99	98	98	99
Propose d method	Bus, Car, Truck	99	99	99	99

VI. CONCLUSION

A frequent challenge in deep learning-based vehicle classification task is obtaining sufficient visual data for the experimentation. In this regard, a customized dataset for vehicle platoon management has been created combining open source repositories and employing image augmentation techniques. The customized dataset has resulted in an imbalanced dataset consisting of 8000 car images, 5000 bus images, and 5000 truck images. The analyses carried out showed that, the performance of classification algorithm with existing methods considering the imbalance dataset is inconsistent across all vehicle classes, lowering the overall performance of the classification task. Whereas, the proposed novel mixed pooling method with three-layer CNN architecture performs significantly well even for the imbalanced dataset. This also highlights the importance of pooling in CNN. Additionally, the selection of the optimal pooling mechanism plays crucial role in optimising the efficiency of the learning algorithm in a CNN architecture. Employing the proposed method, experimental results show that the suggested strategy outperforms the traditional max and mixed pooling approaches by 2% and 8%, respectively for the imbalanced bus and truck datasets. From the accuracy gained using the proposed method, it is evident that selecting the optimal pooling mechanism is crucial for boosting the performance of CNN architecture. Thus, proposed mixed pooling method outperformed other methods on the imbalance dataset. Further research will focus on the performance of the proposed mixed pooling technique with additional CNN layers and a larger number of epochs.

REFERENCES

[1] Wang, F., Dai, H., Lu, Y., and Han, H., "Management and Control of Connected and Automated Vehicle Platoon in the Process of Variable Speed Driving, Vehicle Cut-Out or Cut-In" , SAE Technical Paper 2020-01-5220, 2020.

[2] Zhang, Y., Bai, Y., Hu, J., and Wang, M., "Control Design, Stability Analysis and Traffic Flow Implications of CACC Systems with Compensation of Communication Delay", Transportation Research Record Journal of the Transportation Research Board 0(0):1-15, 202.

[3] C. Peng, M. M. Bonsangue and Z. Xu "Model Checking Longitudinal Vehicle Platoon Systems", in IEEE Access, vol. 7, pp. 112015-112025, 2019.

[4] Venkatesh Muravaneni, Amit Babalal Nahar, and Pratima Vishwakarma, "A Modular Approach to Vehicle Management in a Platoon Group", SAE Technical Paper 2021-26-0125, 2021.

[5] Ajitha; Jeyakumar. S; Yadhu Nandha Krishna K; A Sivasangari, "Vehicle Model Classification Using Deep Learning", IEEE 5th International Conference on Trends in Electronics and Informatics (ICOEI), June 2021.

[6] M. A. Hedeya, A. H. Eid and R. F. Abdel-Kader, "A Super-Learner Ensemble of Deep Networks for Vehicle-Type Classification", IEEE Access, vol. 8, pp. 98266-98280, 2020.

[7] Song, H., Liang, H., Li, H. et al "Vision-based vehicle detection and counting system using deep learning in highway scenes", Eur. Transp. Res. Rev. 11, 51 (2019).

[8] V Sowmya, R Radha, "Real-time Vehicle Detection implementing Deep Convolutional Neural Network Features Data Augmentation Technique", Indian Journal of Science and Technology 2021;15(1):44-5.

[9] Y. Lou, Y. Bai, J. Liu, S. Wang, and L. Duan," VERI-wild: A large dataset and a new method for vehicle re-identification in the wild", Proc. EEE/CVF Conf. Computer Vis. Pattern Recognition. (CPR), Jun. 2019, pp. 3235_3243.

[10] M. Naphade, S. Wang, D. C. Anastasiu, Z. Tang, M.C. Chang, X. Yang, Y. Yao, L. Zheng, P. Chakraborty, C. E. Lopez, and A. Shama " The 5th AI city challenge", Proc. IEEE/CVF Conf. Computer. Vis. Pattern Recognition Workshops, Jun. 2021.

[11] T. Ki and Y. Hur, "Deep Scattering Network with Max-Pooling", Data Compression Conference (DCC), 2021, pp. 348-348.

[12] T. Theodoridis, K. Lumponias, N. Vretos, and P. Daras, "Zernike Pooling: Generalizing Average Pooling Using Zernike Moments", IEEE Access, vol. 9, pp. 121128-121136, 2021.

[13] S. Y. Jo, N. Ahn, Y. Lee, and S. J. Kang, "Transfer learning-based vehicle classification," Proceedings of the 2018 International SoC Design Conference (ISOCC), pp. 127-128, IEEE, Daegu, South Korea, 2018.

[14] Jonathan Krause, Michael Stark, Jia Deng, Li Fei-Fei, "3D Object Representations for Fine-Grained Categorization", 4th IEEE Workshop on 3D Representation and Recognition, at ICCV(3dRR-13),Sydney, Australia,2013.

[15] Tabassum, Shaira, Ullah, Md. Sabbir, Al-nur, Nakib Hossain, Shatabda, Swakkhar 2020. "Poribohon-BD". Mendeley Data, V2.

[16] https://github.com/datacluster-labs/Indian-Vehicle-Image-Dataset.

[17] Tsai, C.C., Tseng, C.K., Tang, H.C., Guo, J. "Vehicle detection and classification based on deep neural network for intelligent transportation application", APSIPA Annual Summit and Conference 2018, IEEE.

[18] He K., Zhang X., Ren S., Sun J., "Spatial pyramid pooling in deep convolutional networks for visual recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 37, 9, 2015, 1904-1916.

[19] Yu D., Wang H., Chen P., Wei Z., "Mixed pooling for convolutional neural networks", International Conference on Rough Sets and Knowledge Technology, 2014, 364-375.

[20] Zeiler M. D., Fergus R., "Stochastic pooling for regularization of deep convolutional neural networks", arXiv preprint arXiv:1301.3557, 2013, 1-9.

[21] Hongye Liu, Yonghong Tian, Yaowei Yang, Lu Pang, and Tiejun Huang, "Deep relative distance learning: Tell the difference between similar vehicles", Proc. CVPR, pages 2167–2175, 2016.

[22] Ke Yan, Yonghong Tian, Yaowei Wang, Wei Zeng, and Tiejun Huang, "Exploiting multi-grain ranking constraints for precisely searching visually-similar vehicles", Proc. ICCV, pages 562–570, 2017.

[23] Chen, L., Ye, F., Ruan, Y. *et al*. "An algorithm for highway vehicle detection based on convolutional neural network." J Image Video Proc. 2018**,** 109 (2018).

[24] X. Wang, W. Zhang, X. Wu, L. Xiao, Y. Qian, and Z. Fang, "Real-time vehicle type classification with deep convolutional neural networks", Journal of Real-Time Image Processing, vol. 16, no. 1, pp. 5–14, 2019.

[25] M. S. Chauhan, A. Singh, M. Khemka, A. Prateek, and R. Sen, "Embedded CNN based vehicle classification and counting in non-laned road traffic", Proceedings of the Tenth International Conference on Information and Communication Technologies and Development, pp. 1–11, Ahmedabad, India, 2019.

[26] Wang, Y., Ban, X., Wang, H., Wu, D., Wang, H., Yang, S., Liu, S., Lai, J, "Detection and classification of moving vehicle from video using multiple spatio-temporal features", recent advances in video coding and security. IEEE Access 7, 80287–80299 (2019).

Framework to Develop a Resilient and Sustainable Integrated Information System for Health Care Applications: A Review

Ayogeboh Epizitone
ICT and Society Research Group,
Information and Corporate Management, Durban University of Technology, Durban, South Africa

***Abstract*—The reconstruction of the health sector amidst the forth industrial revolution has been confronted with many challenges. Many benefits have been attributed to the vital role played by technology in realizing and constructing a robust health information system. However, amidst the digitalization in the healthcare system, several challenges such as integration and fragmentation have been affecting the structure of the Health Information Systems (HIS) which subsequently influences decision making and resource allocation. Therefore, this paper through a comprehensive systematic review afford a proposition for a develop a resilient and sustainable information system for Health Care applications. The study reveals the parallel impact of health information technology application in the healthcare arena and highlight the need for more in-depth research on HIS that incorporate novel scientific methods. Additional this study also presents a body of evident that reveal the inadequacies of the HIS to tackle the constant transformative changes presently confronting the global healthcare systems.**

***Keywords*—Health information system; integrated information system; e-health; bioinformatics**

I. Introduction

Over the last decades there have been remarkable advancements in medical informatics with the development of the health information systems (HIS) championing this course [1, 2]. This has benefited the continuous changing societies in the attainment of sustainable health care as well as efficient and effective information management. An increase in life expectancy and resource allocation are among some of the known attributes foster by this progress. The current HIS developmental path has been gearing toward a global information system to cater for a universal health coverage and civil registration and vital statistics [3]. However, despite the unique contribution of the HIS in shaping health care systems and applications, there are still many current challenges associated with the implementation of the HIS.

Theoretically, the HIS is commended as a system that plays an important role in the integration of different departments as well as support information flow [4]. Practically these systems are considered to be asymmetrical, resulting to distorted functions within the health care environment [3]. Highlighting the need for efficient development and strategic management of HIS. However, despite the strong influence exerted by Information and Communication Technology (ICT) in the realizations of the diverse health care goals [5]. Previous research on information systems has been reported to be short of due diligence in tackling the global healthcare systems transformation [6]. Hence, the need for study that employ novel scientific approaches such as an integrated data science approach and machine learning techniques is indispensable and essential for development of a framework that enhance the current HIS, as well as providing practical findings to realigns the healthcare global transformation.

Over that last decades, information technology (IT) has presented several opportunities as well as equal challenges to many disciplines. In fact, within the healthcare arena, IT through the HISs has facilitated many benefits and transformations such as eHealth [6] and apt decision making that has attracted many nations to invest and assimilate these systems [2, 6, 7]. However, amidst this digitalization in the healthcare arena, there have been several challenges presented such as integration, data quality and the structure of the HIS [3, 6, 8, 9]. Authors also highlight the lack of concerns on the transformative changes in the Healthcare industries and the need for a strengthened HIS [3, 5-7]. Therefore, studies that tackles these problems are vital and necessary for the strengthening of a nations' HIS. This paper seeks to explore extant literature base on HIS to ascertain the knowledge, constraints and perceptions on Health Information Systems deliverables, and afford a proposition for resilient and sustainable HIS.

II. Literature Review

The contributions of an integrated information system in the healthcare arena have been alluded in literature with several authors attesting to the value creation it brings to the field [6, 8, 10-12]. According to English, Masilela, Barron and Schonfeldt [4] a robust integrated information system is pivotal to an efficacious healthcare delivery. Many attributes have been accorded to the successful deployment of HIS especially in the current times [5, 6, 8, 13]. However, there are many persistent challenges revolving around the technological and institutional scopes [2, 4-7, 13, 14].

Research on HIS is considered to be of great importance due to the role it plays in realizing the two urgent priorities of globally health; universal health coverage, and civil registration and vital statistics systems [3, 15, 16]. Similarly, HIS's research is also asserted to be pivotal in supporting the attainment of sustainable development goals by 2030 [1, 3, 5, 17]. However, despite this disposition, a large amount of research done on HIS recurrently highlight major shortcomings in prior studies, in meeting the developmental and anticipated transformation confronting the complex healthcare system [3, 6, 7]. Ostern, Perscheid, Reelitz and Moormann [6],criticize HIS research for focusing on trifling agenda such as, investment and failing to assimilate healthcare industries transformative concerns. Literature further reveal the unsatisfactory trepidations from scholars on addressing potential long lasting changes in the healthcare industries [6, 7].

Critical review of literature highlights the need for well-prepared HIS research to exploit it potentially to tackle healthcare systematic problems thoroughly. In order to afford diverse solutions apt for the healthcare industries and its many actors such as the healthcare providers, government and technology companies [6]. Lluch [11], alluded the need for further research for HIS optimal development and cost effective applications. Scholars, further indicated the deficiency of adequate HIS [5, 8]. Najimudeen, Aldheleai and Ubaidullah [8] identified the impeded widespread use of HIS to be attributed to implementation challenges such as lack of; basic facilities, experts, medical personals e-readiness and technical complexities. Dehnavieh, Haghdoost, Khosravi, Hoseinabadi, Rahimi, Poursheikhali, Khajehpour, Khajeh, Mirshekari and Hasani [12], categories these challenges as operational basing them on the experiences of the actors in the healthcare. These authors call for more research to strengthen the HIS [4, 5, 12].

Literature also revealed many HIS research to be qualitative lacking empirical finding and scientific rigour [3, 5, 6]. Despite the growing data and advancement of ICTs in the health sector, Authors purported that there is a distinct dearth of macro analysis that elucidate impact [18]. Suresh and Singh [18], argued that impact data for ICTs in health development remain a grand challenge.

Many studies also reveal the challenges faced by developing countries implementing HIS [8, 18]. While many advances have been attained by developed nation, Dehnavieh, Haghdoost, Khosravi, Hoseinabadi, Rahimi, Poursheikhali, Khajehpour, Khajeh, Mirshekari and Hasani [12] reveal the tussle in developing countries where several barriers hinder the development of HIS which resultantly influences decision making due to the chaotic and fragmented state of HIS [18, 19]. These authors allude the need for developing countries to pay attention to the local needs, priority, environment, infrastructure and capacity to harness technological solutions.

Additional, many studies have been conducted in developed countries like Australia, USA and Europe and practical solutions and framework for HIS have developed, implemented and adopted [18]. Whereas in developing countries there is a deficit in studies and solutions on HIS phenomena [12, 18]. Some studies have suggested that developing countries harbor many barriers to HIS development such as inadequate resources, lack of data ownership and feedback [12, 18, 19].

III. Methodology

The important of HIS has been herald in literature with the challenges hindering it implementation and optimization. The demand for a better understanding of the phenomena confronting HIS development, utilization and optimization is paramount to the effectiveness of these systems in the places where they are deployed. Couple with the necessities of the global priorities and sustainable goals for a global healthcare system. The need to comprehensive review extensive literature is paramount and a needed step in the attainment of a sustainable and resilient HIS. This study being qualitative in nature and aiming to uncover insight on HIS utilizes a comprehensive systematic review [14, 20]; adopting an abductive approach that ensue a systematic content analysis procedure on extant literature and studies of HIS research [20-23].

Preceding the review, search engines are utilized to perform search on online databases and websites using the keywords to retrieve the needed data for this study. For the purpose of adequately responding to the study objective to uncover insight that aid HIS sustainability and resilient, articles containing keywords and references to “Health Information System” AND “HIS” are analysis painstaking. Identification of these articles was done through an exhaustive literature search on prominent journals and database feature studies on HIS. Table I outlines some of the journals, databases and Keywords used.

TABLE I. Information Source for HIS journal

Information Sources	
Journals and Databases	Keywords
Web of Sciences	Health information system
Google Scholar	Integrated Information system
Elsevier	E- Health
ProQuest	Mobile health
Science Direct	
PubMed	
Embase	

IV. Discussion and Result

Literature reveal the urgent parallel need for global health information systems that afford valuable aid to effective decision making which is paramount in rendering effective healthcare applications [18, 24]. The review on health information system was conducted in line with prior methodology that explores extant literature extensively to ascertain current insight and knowledge of these systems.

A. ICT and HIS

Many facets of the HealthCare segment have been significantly impacted by ICT [18]. Making the value of an effective information system indispensable in today era. Literature attribute the difference between good and bad decision to be associated with the difference between life and death [18]. Authors posit the optimization of ICT capabilities in the healthcare arena can be realize via the fusion of primeval and novel technologies [18]. According to Suresh and Singh

[18], ICT serve as an irreplaceable mediator for community access to Healthcare and information. Highlighting the detrimental role of information in effective decision-making. Similarly, Najimudeen, Aldheleai and Ubaidullah [8], revere ICT mediation in facilitating access to medical care and information, branding its support within the healthcare sector as "healing at a distance". However, despite the involvement of ICT in improving healthcare, several authors have advocated for the redesigning of ICT within the healthcare arena [14, 25].

B. Benefits of ICT in HealthCare Delivery

The appraisal of integrated information systems like HIS has always been center on the growing numbers of implementation. However, many studies have recounted that increase implementations and adoptions does not equate success. Literature, places emphases on improvement, efficiencies and effectives of outcomes and methods directly concerns with healthcare as a true definition of success [18]. Suresh and Singh [18] reports the important of health information that serve as vehicle to track health needs, guide health programs design and implementation and quality assessment. A similar study by Najimudeen, Aldheleai and Ubaidullah [8] enumerate on the value of HIS in health care delivery. According to Najimudeen, Aldheleai and Ubaidullah [8], HIS has afforded tremendous opportunities that includes the plunging of medical errors, advancing of healthcare admin efficiency and medical information management.

C. Challenges in Health Information Systems

Despite the advancement and development of the ICT applications in the health sectors there exist several challenges. Some authors have attributed these challenges to be foster by issues such as connectivity, ICT literacy and lack of technological convergence [14, 18, 24]. According to Fusheini and Eyles [24] the effectiveness and efficiency of healthcare delivery is significantly affected by these challenges. Contemporary studies highlight the lack of incorporation of stakeholder and technological design flaw to be contributing to these challenges[14, 25]. A study by Dehnavieh, Haghdoost, Khosravi, Hoseinabadi, Rahimi, Poursheikhali, Khajehpour, Khajeh, Mirshekari and Hasani [12], categorizes these challenges in themes that included human resources, data issues and infrastructure among many others. The Table II summarizes these challenges identified in the diverse studies.

Funding has been identified in several studies to be a critical challenge influencing the implementation and strengthening of HIS. Dehnavieh, Haghdoost, Khosravi, Hoseinabadi, Rahimi, Poursheikhali, Khajehpour, Khajeh, Mirshekari and Hasani [12], contend that the effectiveness and optimization of HIS is hinder and influence by financial deficiencies. Fusheini and Eyles [24] added to this claim revealing the lack of ample resources to afford funding. Many authors highlighted the need for the inclusion of financial needs associated with HIS in the annual plans [12, 26].A similarly study by Bergum, Kusumasindra, Øren, Falch and Sahraoui [27] and Al-Nashy [28] support this assertion and emphasis on the need for a sustainable finance for HIS implementation. These authors indicate the vitality of stable and sufficient financial sources for HIS enactment. Implying the lack of acknowledgement and incorporation of funding for HIS, results to shortage of finances needed to address financial related needs associated with HIS function and implementation as they arise.

TABLE II. EXTANT HIS CHALLENGES

Challenges	**Cited Sources**
Finance (Funding Sources)	Manoj et al. 2013; Al-Nashy 2015; Bergum et al. 2015; Fusheini and Eyles 2016; Dehnavieh et al. 2019.
Communication Infrastructure	Dehnavieh et al. 2019; Najimudeen, Aldheleai and Ubaidullah 2021.
Data (completeness, ownership and security)	Kiberu et al. 2014; Karuri et al. 2014; Suresh and Singh 2014; Dehnavieh et al. 2019; Farnham et al. 2020.
Political, cultural, social and structural infrastructure	Suresh and Singh 2014; Fusheini and Eyles 2016; Dehnavieh et al. 2019.
Workforce capacity	Manoj *et al.* 2013; Karuri *et al.* 2014; Kiwanuka, Kimaro and Senyoni 2015; Dehnavieh *et al.* 2019.
Top Management and leadership	Sheikh and Bakar 2011; Manya et al. 2012; Adaletey, Poppe and Braa 2013; Manoj et al. 2013; Karuri et al. 2014;Manya and Nielsen 2015; Dehnavieh et al. 2019.
Lack of Training	Sheikh and Bakar 2011; Adaletey, Poppe and Braa 2013; Manoj et al. 2013; Karuri et al. 2014; Al-Nashy 2015; Kiwanuka, Kimaro and Senyoni 2015; Manya and Nielsen 2015; Nguyen 2015; Dehnavieh et al. 2019.
Project Management	Sheikh and Bakar 2011; Manoj et al. 2013; Dehnavieh et al. 2019
Application selection criteria	Dehnavieh *et al.* 2019
Stakeholder coordination.	Dehnavieh *et al.* 2019; Grosjean, Bate and Mestre, 2020; Grosjean et al., 2022
Pre-deployment (Pilot System).	Sheikh and Bakar 2011; Manoj et al. 2013; Al-Nashy 2015; Dehnavieh *et al.* 2019

Communication Infrastructure which comprises of ICTs' infrastructure for Internet, Mobile and Electricity's infrastructure has been in the center of Healthcare sector. These infrastructures are vital to the efficient and effective functioning of operations in the healthcare arena. However, several studies have reported communication infrastructure to be inapt [8, 12]. Authors argue that the inadequacies of supporting infrastructure are among the challenges disclosed by stakeholders [8]. Several studies highlight the internet connection restriction in many countries to be problematic to the utilization of HIS in many areas [12].

Data management has influence several aspects in the health sector and many decision-making has been reliance on quality and analysis of data [29]. However, there is an information usage deprivation for decision making in the healthcare arena [12, 30]. Even though information and data usage has been encouraged and developed in many countries, it still remains an enormous problem. According to Karuri, Waiganjo, Daniel and Manya [19], data management has significantly influence health decision-making and funding allocation [12, 18, 19]; indicating data management to be a significant determinant in the optimization and implementation for HIS. Dehnavieh, Haghdoost, Khosravi, Hoseinabadi, Rahimi, Poursheikhali, Khajehpour, Khajeh, Mirshekari and Hasani [12] argue that all countries encounter data related problems with some experiencing data concerns more than the others.

Infrastructures are a formidable foundation in many HIS deployment. Literature reveals the challenges presented by infrastructures like political, cultural, social and structural to the implementation of HIS to be daunting. These infrastructures are considered to serve as a major challenge to HIS with many citations in literature [12, 24]. Hence, some authors recon they be ascertained forthright in relation to the HIS efficacy [12, 18]. Suresh and Singh [18] indicate their availability and adequacy to be paramount to HIS. Whilst, Dehnavieh, Haghdoost, Khosravi, Hoseinabadi, Rahimi, Poursheikhali, Khajehpour, Khajeh, Mirshekari and Hasani [12], further reveal several challenges such as political instability, linguistic, regional cooperation and technical-economic associated with these infrastructures that indirectly or directly influence HIS operation. Additionally, Fusheini and Eyles [24] reveal the adverse outcome of this constraint on healthcare delivery.

Workforce capacity are also among the principal challenges face by the HIS implementation and optimization. Many study reveal the efficacy of HIS to be reliance on the human competency and capacity of staff [19, 31]. However, many studies allude workforce shortage, incompetency, accuracy and motivation as dares confronting HIS efficacy [12, 19, 26].

Top Management and leadership has been herald in many information systems to be a dare that hinder the effectiveness and efficacies of these systems. Likewise, in the HIS, several top management and leadership functions have been reported to be vital in the operation of HIS. Among which are planning, participation, perception and support that are alleged to be inadequate [12, 26]. Authors highlight the need to win top management from all level that includes, national and provincial levels for HIS implementation [19, 26]. Similarly, Adaletey, Poppe and Braa [32] and Manya, Braa, Øverland, Titlestad, Mumo and Nzioka [33] cheer the attainment of these key stakeholders involvement in HIS deployment. Many other studies call for the restructuring of management role and early formation of team [12, 19, 34, 35].

Deficiency of Training has been professed in literature to significantly reduce the HIS optimization [12, 19, 26, 34, 35]. Many studies call for the need to provide training to stakeholders, highlighting their lack of skills and competencies to be attributed to the absence of adequate trainings [10, 12, 19, 26, 28, 31, 32]. Study done by Manya and Nielsen [34] and [26], accentuate the need for training methods that include workshops and formal educational courses that incorporate biomedical informatics course.

Other challenges found in literature include project management, application selection criteria, stakeholder communication and coordination and pre-deployment (Pilot System) [12, 26, 35]. Their inadequacy serves as a major barrier to HIS implementation and efficacy.

V. Conclusion and Future Research

The importance of HIS and it capabilities in the delivery and deployment of quality healthcare application cannot be understated. As the call for an improve and universal healthcare advances, it is pertinent to explore and develop effective models to attain these objectives. Although many developing countries have already enacted technological solutions to enhance their healthcare section. The findings of this study show that there exists a strong body of evidence in literature that highlight the unfitness of the HIS for the transformative changes confronting the globe today. Correspondingly, several studies reveal the need for a strong and hybrid HIS to eliminate some of the present challenges confronting the deployment of an integrated information system like the District HIS (DHIS).

Notwithstanding the commendable benefits of the HIS, the findings reveal there is a need and call for active engagements of countries to strengthening these systems in alignment with the global priorities. While many studies have been undertaken in regards to HIS, the plethora of challenges confronting its implementations and utilization seem to remain a grand huddle in the advancement of healthcare applications. Eleven of these challenges was reported in this study with seven most prominent discussed. The challenges reveal the inadequacies of the HIS to tackle the constant transformative changes presently confronting the global healthcare systems.

Additionally, the findings of this study highlights the necessity for further research on HIS deployment optimization that take into consideration their challenges and concerns. And also, provide valuable insight into HIS offerings and dares for decision maker and HIS stakeholders.

Acknowledgment

The author would like to thank the Durban University of Technology writing retreat committee for their support and assistance.

References

[1] R. Haux, “Health information systems–past, present, future,” International journal of medical informatics, vol. 75, no. 3-4, pp. 268-281, 2006.

[2] H.-A. Park, “Are we ready for the fourth industrial revolution?,” Yearbook of medical informatics, vol. 25, no. 01, pp. 1-3, 2016.

[3] S. Sahay, P. Nielsen, and M. Latifov, “Grand challenges of public health: How can health information systems support facing them?,” Health policy and technology, vol. 7, no. 1, pp. 81-87, 2018.

[4] R. English, T. Masilela, P. Barron, and A. Schonfeldt, “Health information systems in South Africa,” South African health review, vol. 2011, no. 1, pp. 81-89, 2011.

[5] S. Sahay, A. Rashidian, and H. V. Doctor, “Challenges and opportunities of using DHIS2 to strengthen health information systems in the Eastern Mediterranean Region: A regional approach,” The Electronic Journal of Information Systems in Developing Countries, vol. 86, no. 1, pp. e12108, 2020.

[6] N. Ostern, G. Perscheid, C. Reelitz, and J. Moormann, “Keeping pace with the healthcare transformation: a literature review and research agenda for a new decade of health information systems research,” Electronic Markets, vol. 31, no. 4, pp. 901-921, 2021.

[7] L. Chen, A. Baird, and D. W. Straub, "An analysis of the evolving intellectual structure of health information systems research in the information systems discipline." Association for Information Systems, 2019.

[8] M. Najimudeen, H. F. Aldheleai, and M. Ubaidullah, “Health Care Professionals' Use of Health Information Systems (HIS) in Indian Hospitals,” International Journal of Computer Applications, vol. 975, pp. 8887, 2021.

[9] L. Nguyen, E. Bellucci, and L. T. Nguyen, “Electronic health records implementation: an evaluation of information system impact and contingency factors,” International journal of medical informatics, vol. 83, no. 11, pp. 779-796, 2014.

[10] S. Nguyen, “User acceptance of instant messaging in DHIS 2,” Halden, Norway: Ostfold university college, pp1-96, 2015.

[11] M. Lluch, “Healthcare professionals’ organisational barriers to health information technologies—A literature review,” International journal of medical informatics, vol. 80, no. 12, pp. 849-862, 2011.

[12] R. Dehnavieh, A. Haghdoost, A. Khosravi, F. Hoseinabadi, H. Rahimi, A. Poursheikhali, N. Khajehpour, Z. Khajeh, N. Mirshekari, and M. Hasani, “The District Health Information System (DHIS2): A literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries,” Health Information Management Journal, vol. 48, no. 2, pp. 62-75, 2019.

[13] T. Kivinen, and J. Lammintakanen, “The success of a management information system in health care–A case study from Finland,” International Journal of Medical Informatics, vol. 82, no. 2, pp. 90-97, 2013.

[14] S. Grosjean, E. Bate, and T. Mestre, “Designing socially acceptable mHealth technologies for Parkinson's disease self-management,” Finnish Journal of eHealth and eWelfare, vol. 12, no. 3, pp. 163-178, 2020.

[15] C. Dye, J. C. Reeder, and R. F. Terry, “Research for Universal Health Coverage,” Science Translational Medicine, vol. 5, no. 199, pp. 199ed13-199ed13, 2013.

[16] WHO, Consultative expert working group on research and development: financing and coordination: Report of regional technical consultation, Bangkok, thailand, 15-17 August 2012, WHO Regional Office for South-East Asia, 2013.

[17] C. AbouZahr, and T. Boerma, “Health information systems: the foundations of public health,” Bulletin of the World Health Organization, vol. 83, pp. 578-583, 2005.

[18] L. Suresh, and S. N. Singh, “Studies in ICT and Health Information System,” International Journal of Information Library and Society, vol. 3, no. 1, pp. 16, 2014.

[19] J. Karuri, P. Waiganjo, O. Daniel, and A. Manya, “DHIS2: the tool to improve health data demand and use in Kenya,” Journal of Health Informatics in Developing Countries, vol. 8, no. 1, 2014.

[20] A. Epizitone, and O. O. Olugbara, “Mixed method approach to determination critical success factors for successful financial ERP system implementation,” Academy of Accounting and Financial Studies Journal, vol. 24, no. 2, pp. 1-10, 2020.

[21] G. R. Dantes, and Z. A. Hasibuan, “Enterprise resource planning implementation framework based on key success factors (KSFs),” UK Academy for Information System, pp. 11-13, 2011.

[22] A. Epizitone, and O. O. Olugbara, “Critical success factors for ERP system implementation to support financial functions,” Academy of Accounting and Financial Studies Journal, vol. 23, no. 6, pp. 1-11, 2019.

[23] B. M. Kalema, O. O. Olugbara, and R. M. Kekwaletswe, “Identifying critical success factors: The case of ERP systems in higher education,” The African Journal of Information Systems, vol. 6, no. 3, pp. 1, 2014.

[24] A. Fusheini, and J. Eyles, “Achieving universal health coverage in South Africa through a district health system approach: conflicting ideologies of health care provision,” BMC Health Services Research, vol. 16, no. 1, pp. 1-11, 2016.

[25] S. Grosjean, J.-L. Ciocca, A. Gauthier-Beaupré, E. Poitras, D. Grimes, and T. Mestre, “Co-designing a digital companion with people living with Parkinson's to support self-care in a personalized way: The eCARE-PD Study,” Digital Health, vol. 8, pp. 20552076221081695, 2022.

[26] S. Manoj, A. Wijekoon, M. Dharmawardhana, D. Wijesuriya, S. Rodrigo, R. Hewapathirana, P. Siribaddana, T. Gunasekera, and V. H. Dissanayake, “Implementation of district health information software 2 (DHIS2) in Sri Lanka,” Sri Lanka Journal of Bio-Medical Informatics, vol. 3, no. 4, 2013.

[27] B.-I. Bergum, F. Kusumasindra, M. Øren, V. Falch , and T. Sahraoui, “Information infrastructure: Deliverable 2 analyzing DHIS2 as an information infrastructure. Submission Date: November 5, 2015. University of Oslo, https://www.uio.no/studier/emner/matnat/ifi/INF5210/h15/project-reports/analysing-dhis2-in-ghana-final-delivery.pdf,” pp. pp. 22., 2015.

[28] S. A. T. Al-Nashy, “Managing scaling of HIS: implementation of DHIS2 in Sudan. Available at: https://www.duo.uio.no/bitstream/handle/10852/43905/1/Al-Nashy-Master.pdf. Accessed 21/03/2022,” 2015.

[29] V. M. Kiberu, J. K. Matovu, F. Makumbi, C. Kyozira, E. Mukooyo, and R. K. Wanyenze, “Strengthening district-based health reporting through the district health management information software system: the Ugandan experience,” BMC medical informatics and decision making, vol. 14, no. 1, pp. 1-9, 2014.

[30] A. Farnham, J. Utzinger, A. V. Kulinkina, and M. S. Winkler, “Using district health information to monitor sustainable development,” Bulletin of the World Health Organization, vol. 98, no. 1, pp. 69, 2020.

[31] A. Kiwanuka, H. C. Kimaro, and W. Senyoni, “Analysis of the acceptance process of district health information systems (DHIS) for vertical health programmes: a case study of TB, HIV/aids and malaria programmes in Tanzania,” The Electronic Journal of Information Systems in Developing Countries, vol. 70, no. 1, pp. 1-14, 2015.

[32] D. L. Adaletey, O. Poppe, and J. Braa, "Cloud computing for development—Improving the health information system in Ghana." In 2013 IST-Africa Conference & Exhibition pp. 1-9. IEEE, 2013.

[33] A. Manya, J. Braa, L. H. Øverland, O. H. Titlestad, J. Mumo, and C. Nzioka, "National roll out of District Health Information Software (DHIS 2) in Kenya, 2011–Central server and Cloud based infrastructure." In 13th international conference on social implications of computers in developing countries Vol. 3. 2015.

[34] A. Manya, and P. Nielsen, "The use of social learning systems in implementing a web-based routine health information system in Kenya." In 13th international conference on social implications of computers in developing countries, vol. 3. 2015.

[35] Y. H. Sheikh, and A. D. Bakar, "Open source software solution for healthcare: the case of health information system in Zanzibar." In International Conference on e-Infrastucture and e-Services for Developing Countries, pp. 146-155. 2011.

TEC Forecasting using Optimized Variational Mode Decomposition and Elman Neural Networks

Maladh Mahmood Shakir, Zalinda Othman, Azuraliza Abu Bakar
Faculty of Information Science and Technology
University Kebangsaan Malaysia
Bangi, Selangor, Malaysia

***Abstract*—Forecasting the ionosphere layer's total electronic content (TEC) is crucial for its impact on satellite signals and global positioning systems (GPS) and the ability to predict earthquakes. The existing statistical-based forecasting models such as ARMA, ARIMA, and HW suffered from the TEC non-stationarity nature, which requires algorithmic handling of the forecasting and the mathematical part. This study proposes a hybrid method that incorporates several components and is designated as Optimized Variational Mode Decomposition with Recursive Neural Network Forecasting (OVMD-RNN) to forecast TEC. Before using the Elman Network to train each component, Variational Mode Decomposition (VMD) was used to decompose the signal into its essential stationary components. In addition, the proposed method includes an optimization algorithm for determining the best VMD decomposer parameters. The GPS Ionospheric Scintillation and TEC Monitor (GISTM) at Universiti Kebangsaan Malaysia station have been used to evaluate the method based on collected datasets for three years, 2011, 2012, and 2013. The experiment findings show that the model has successfully tracked all the up and down patterns in the time series. The results also reveal that VMD-based training might not always provide good results due to the residual signal. Finally, the evaluation focused on generating loss value and comparing it to the ARIMA benchmark. It showed that OVMD-RNN had accomplished a maximum improvement percentage of ARIMA with a value of (99%).**

***Keywords*—Elman neural networks; forecast; hybrid model; optimized Variational Mode Decomposition; total electronic content**

I. Introduction

The global navigation satellite system (GNSS) has become a critical system for providing a wide range of services and applications in the modern world. As a result, its dependability and performance are critical for various systems. The state of the ionosphere and its amount of influence by solar radiation and the geomagnetic field is one of the factors that affect GNSS radio communication transmissions [1].

Researchers utilise a quantitative metric to explain the effect of solar radiation on the total electron content (TEC) of the ionosphere layer. This variable represents the variability of solar radiation's impacts on the ionosphere layer, as well as it's geographically and temporally represented global positioning system (GPS) coordinates and time information. A dependable and accurate TEC forecasting can provide useful feedback to GPS receivers and improve numerous GPS-dependent services. Scientists created the International Reference Ionosphere (IRI) project to anticipate electron density, ion composition, electron and ion temperature, and vertical electron column density due to the relevance of ionosphere prediction. As a result, IRI is a collaborative international initiative of the Committee on Space Research (COSPAR) and the International Union of Radio-science (URSI) to build and refine a reference model for the Earth ionosphere's most critical plasma properties [2].

The main goal of majoring in TEC is to foresee any delays in GPS or communication signals in general, which could impair the operation of numerous devices. Earthquake forecasting is another application of measuring and evaluating TEC. Studies have shown that the ionosphere is influenced by the Sun's position and radiation and geomagnetic activity, and seismic activity in the Earth's crust and surface. These impacts can aid in predicting an earthquake several days ahead of time [3].

The challenging aspect of TEC prediction is its non-stationarity nature. This has brought a limitation to the existing studies that uses statistical-based forecasting models such as ARMA, ARIMA, and HW [4]. Hence, researchers have tested the capability of neural networks in the general and recursive types of neural networks [5]. The latter is more preferred because of its feedback or memory aspect, which enables more flexibility in modelling the dynamics of the time series. However, RNN is not sufficient because of embedded non-stationarity in the time series, which requires algorithmic handling of the forecasting in addition to the mathematical part. VMD decomposition is a possible candidate for decomposing the original time series into various stationary components [6], However, it is still encountering an issue in deciding the optimal number of intrinsic mode components for better forecasting performance and obtaining optimal VMD components before forecasting is needed. Hybrid architecture is a good candidate approach through combining RNN with optimal VMD settings. With such a hybrid combination, the VMD will decompose the original time series into stationary components that are easier to be learned by the following RNN layer.

The current forecasting models suffered from the TEC non-stationarity nature. This study describes a new method for forecasting the TEC time series that combines three components: decomposing the original signal to its stationary components using Variational Mode Decomposition (VMD), parameter optimization using Genetic Algorithm (GA), and forecasting using the Elman Neural Network, a well-known Recurrent Neural Network (Elman NN).

In the next Section II, related works on TEC forecasting and current approaches have been presented. In Section III, the proposed method has been described, i.e., OVMD-RNN and the methodology flow. The experimental works have been presented in Section IV as well as the results, whereas Section V gives conclusions.

II. Related Work

TEC forecasting increases the knowledge of ionosphere space weather to give accurate warning and mitigation of TEC impacts. Due to the time-series nature of TEC forecasting, researchers prefer traditional time series forecasting models such as Auto Regression Moving Average (ARMA) and its derivatives [7]. The VMD-ARMA (VARMA) model is described in [8] as a non-stationary signal decomposition technique based on Variational Mode Decomposition (VMD) paired with Auto Regressive Moving Average (ARMA) to estimate ionosphere delay values 1-hour ahead.

However, despite their large training requirements, some researchers have favoured non-linear models with a high approximation, such as Artificial Neural Networks (ANN) [9]. The TEC signal’s non-linear nature is one of the factors driving the researchers' interest in employing ANN models for TEC forecasting. [10] is one of the first researchers to use ANN for TEC forecasting, further aided by [11]. According to them, the capacity to use ANN to predict TEC in places where appropriate training data was absent has been demonstrated empirically. [11] used ANN with a single hidden layer in addition to input and output layers, as described by the 6:9:1 configuration.

According to the newest research, a single hidden layer is sufficient, and adding more layers necessitates more training without increasing accuracy. However, studies have shown that ANN does not outperform basic linear models like ARMA and Auto Regressive Integrated Moving Average (ARIMA) [1], which is essentially ARMA plus an integration component. Both ARMA and ARIMA have outperformed the IRI global reference model. ARMA's and ARIMA's superiority Holt-Winter models, which add a component for seasonable effect, were also seen compared to the global IRI model [12, 13]. Several scholars have made a comparison of ARIMA with ANN. Based on mean absolute error (MAE) and root mean square error (RMSE) values, the performance of the ARMA and ANN models are validated on both geomagnetic quiet and disturbed days [14]. For the ARMA and ANN models, forecasting errors are higher on geomagnetically disturbed days; when tested using MAE and RMSE, ANN outperformed ARMA. Hybrid models of ARMA and ARIMA were used in other investigations. For 1-hour ahead forecast of ionosphere TEC, [12] used hybrid ARIMA models based on Wavelet Transform (WT) and Empirical Mode Decomposition (EMD).

A hybrid GA and ANN model was suggested by [15] to forecast 1-hour vertical TEC for one station in China. The Backpropagation Neural Network (BP) and the GA approach train parameters in a two-step process. GA tunes the weights and biases of the original neural network in the first stage. Model's weights and biases are recorded as a lengthy chromosome. The fitness function expressed as a prediction error is used to evaluate each chromosome's performance in the population. A Wavelet Neural Network (WNN) is utilised to model the ionosphere time series in Iran by [16]. WNN is a hybrid of wavelet theory and neural network theory. One of the advantages of TEC modelling with WNN is its ease of use and speed of computation.

More recent studies have used machine learning algorithms for GPS TEC forecasting. The performance of a Gaussian kernel-based machine learning algorithm was compared to that of ANN and ARMA in [1]. The outcomes of their work have demonstrated dominance over them.

the authors in [17] employed a hybrid model in which the TEC time series was divided into its stationary components using VMD. Then a kernel extreme learning machine was used to anticipate the data. Kernel extreme learning machines were compared to ANN in this study, and the results show that the former is superior in forecasting accuracy. The idea of combining VMD with neural networks can be found in older time series forecasting research. Integration of VMD and generic regression neural networks is the focus of [17].

Another example is [18], who used VMD in conjunction with an extreme learning machine to model the Monthly Precipitation Time Series. Overall, researchers have attempted to forecast TEC time series in various locations for accomplishing needed actions such as handling delays of satellite signals GPS and forecasting earthquakes.

III. Proposed method

This study proposes a time series forecasting model using a VMD, and ENN hybrid architecture called an Optimized Variational Mode Decomposition with Elman Recursive Neural Network Forecasting (OVMD-RNN). It contains the following contributions.

- It develops a framework for TEC forecasting based on decomposition, optimization and ENN.
- It provides GA for finding the optimal IMF components K and the correlation coefficients α for VMD.
- It integrates the developed OVMD by GA with ENN for forecasting the TEC.
- It evaluates the hybrid framework based on our data and compares it with recent models in the literature.

A. Preliminary

1) Variational mode decomposition: The application of VMD is the initial step in the approach. The purpose of VMD is to decompose the TEC signal into bandwidth-limited components of varying frequencies. The decomposition helps eliminate the signal's random behavior and makes it more forecastable. Before VMD, researchers discovered various decomposition techniques, including Fourier transforms, wavelet transforms, and EMD [19]. However, research has shown that VMD outperforms EMD.

In the VMD algorithm [20], the real-valued input signal is viewed as an ensemble of sub-signals (modes) with a narrow band of frequencies around a few light frequencies. The VMD algorithm is represented as a variational problem with constraints:

$$\min_{\{u_k\},\{w_k\}} \left\{\sum_k \left\|\partial t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-jw_kt}\right\|_2^2\right\} \quad (1)$$

$$\text{s.t } \sum_k u_k = f$$

where, u_k denotes variational model components, and w_k are the corresponding centres of each variation model component.

The optimization is solved using Lagrangian multiplier λ and quadratic penalty α as follows;

$$L(u_k,\omega_k,\lambda) =$$

$$\alpha\sum_k \left\|\partial_t\left[\left(\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right)*e^{-j\omega t}\right]\right\|_2^2 \quad (2)$$

$$+\|f-\sum_k u_k\|_2^2+\langle\lambda, f-\sum_k u_k\rangle$$

The augmented Lagrangian equation is then solved using the algorithm of alternate direction multipliers [20].

The values of the number of components K and penalty α play a significant role in VMD decomposition performance [21]. Specifically, a large value of K indicates interferential decomposition, whereas a small value indicates incomplete decomposition. As a result, the best values for K and α must be determined. Therefore. The problem formulation is being reformulated as follows.

Assume y(t) represents a TEC time series. The goal of this study is to predict its value for the future time horizon, Tf. Because of its non-stationary aspect, first, it must be decomposed using the type VMD procedure before forecasting. The result of the decomposition is K modes that are expressed as follows:

$$\sum_{k=1}^{K} u_k(t) = \hat{y} \quad (3)$$

The decomposition's objective function is formulated to maximizes the correlation between intrinsic components Corr such as.

$$Corr = \frac{Cov(y(t),\hat{y}\,(\mathrm{t}))}{\sigma(y(t))\sigma(\hat{y}(t))} \quad (4)$$

$$\begin{cases} [K,\alpha]^* = \\ argmin(1-Corr) = argmin(1-\frac{Cov(y(t),\hat{y}\,(\mathrm{t}))}{\sigma(y(t))\sigma(\hat{y}(t))}) \end{cases} \quad (5)$$

2) Genetic algorithm: A genetic algorithm (GA) is a stochastic searching algorithm with heuristic knowledge. It provides a way to find the optimal value of an objective function based on random generating of candidate solutions, heuristic interaction between them, and selecting elites from one generation to provide the offspring representing the next generation until convergence or meeting the stopping criterion [22]. GA is inspired by the theory of survival of the fittest that was proposed by Darwin [23]. The GA pseudocode is given in Algorithm 1.

Algorithm 1 GA Optimization

```
Input
    S //number of solutions in generation
    N//number of generations
    Objective function
Output
    Best solution
Start
1-  Initiate first population
2-  Current generation =first population
3-  Evaluate current generation
4-  Select elites (using roulette wheel)
1-  Perform crossover ((using uniform crossover) and mutation
    (using probability)
2-  Combine solutions
3-  if not meeting stopping criterion go to 3
4-  best solution =solution of best fitness value of the last
    generation
End
```

GA was employed to optimize the objective function of the decomposition Equation 5, and the two variables K and α represent the chromosome. Algorithm 1 shows the steps of GA to obtain the best K and α values. The first step consists of randomly generating the initial population in this work. The population is evaluated based on the objective function stated in Equation 5. Then, a roulette wheel selection mechanism is applied to select the parent that will undergo the crossover operation, used within a specific probability.

Similarly, the mutation operator is performed on the new solution within a pre-defined probability to maintain the diversity of the population. After that, the produced population will replace the worst solutions of the previous generation. These steps are repeated till meeting the stopping criterion. Finally, GA will return the best solution (i.e., best K and α values).

3) The Elman neural network: An Artificial Neural Network, or ANN, is a massively parallel distributed processing system made up of densely interconnected neural computing parts that can learn, gain knowledge and make it available for use. ANN architecture is defined by the network of neuron connections, the training or learning mechanism for calculating the connection weights, and the activation function.

Even though the Multilayer Perceptron neural network (MLP) can solve a wide range of complex issues, it can only map the input space to the output space in a static way. Elman Neural Network (ENN) [24] is a simple recurrent neural network with dynamic characteristics. The structure of Elman RNN is identical to a three-layered MLP, except for an additional layer called a context layer, which is engaged in the former. In Elman RNN, the hidden neurons are activated by both the input and context neurons. The hidden neurons feed forward to activate the output neurons while also feeding back to activate the context neurons. As a result, the context layer allows ENN to respond to dynamism.

The ENN is trained using the BP algorithm, as shown in Equation 6 [25].

$$W(t+1) = W(t) - \mu\frac{\partial E(t)}{\partial W(t)} \quad (6)$$

where, μ denotes the learning rate.

A conceptual diagram of Elman RNN is presented in Fig. 1.

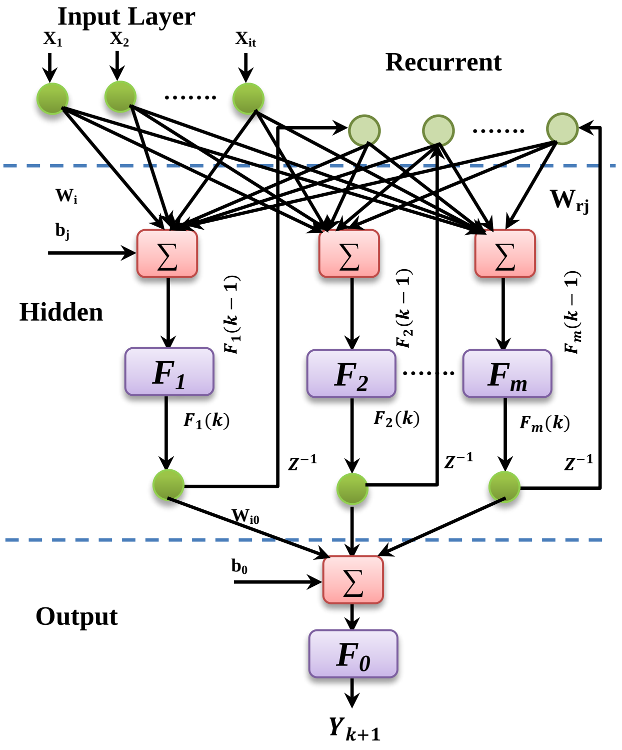


Fig. 1. The Structure of ENN [26].

B. Methodology Flow

As shown in Fig. 2, the developed methodology of the forecasting method consists of three phases after pre-processing: 1- Optimized VMD Decomposition, 2 - Elman RNN training, and 3 - forecasting. The phases are conceptual, but practically they have an interconnected nature. The optimization requires calculating the loss function to rank candidate solutions. On the other hand, the optimal evaluation setting is used for the VMD decomposition. Next, the result is used for ENN training and the trained ENN is used for forecasting.

1) Phase 1: Optimized VMD Decomposition: As is presented in Fig. 2, in the first phase, the data is entered into the optimized VMD decomposer, which is responsible for dividing the time series into various IMF components with the assistance of a genetic algorithm that optimizes K and α values.

2) Phase 2: ENN Training: After decomposing the TEC time series into its IMF components in the first phase, the IMF passed to the recursive neural networks (ENN) to train the ENN model. The training phase comprises of two parts: the first part is the backward path that takes an existing neural network topology and calculates weight changes based on the gradient of the error. The second part is the forward path that uses current weights to propagate.

3) Phase 3 Forecasting: Once the ENN is trained, the testing phase (forecasting) is used. It uses the optimized VMD and trained ENN for forecasting the IMFs and after that a summation of the forecasted time series is used to forecast the overall time series. Such an evaluation includes providing the discrepancies between the predicted values and the actual values.

C. Training and Testing Flow

The interaction between the training phase and the testing phase is given in Fig. 3. As it is shown in the figure, the data comes as s stream, and it is partitioned into two parts: the historical part w_h and the future part w_f. The data in w_h was used for training, while the data in w_f was for performance evaluation. For the testing, the predicted value was compared with the original value stored in w_f and used to calculate testing errors.

D. Data pre-processing and partitioning

Pre-processing comprises data visualization and determining whether missing records exist. In the case of missing records, an average window over multiple days is used to replace the missing value in one day. After that, the data were partitioned. The purpose is to divide the data into training and testing data. The training data will come from the past, whereas the testing will occur in the future, forecasting time intervals. The signal and the duration of the time window are inputs. The output is a matrix consisting of training and testing data. The data will be passed through the time window, and samples will be added. The final sample is a prediction based on these samples. Following data collection, the data is divided into two categories: training and testing.

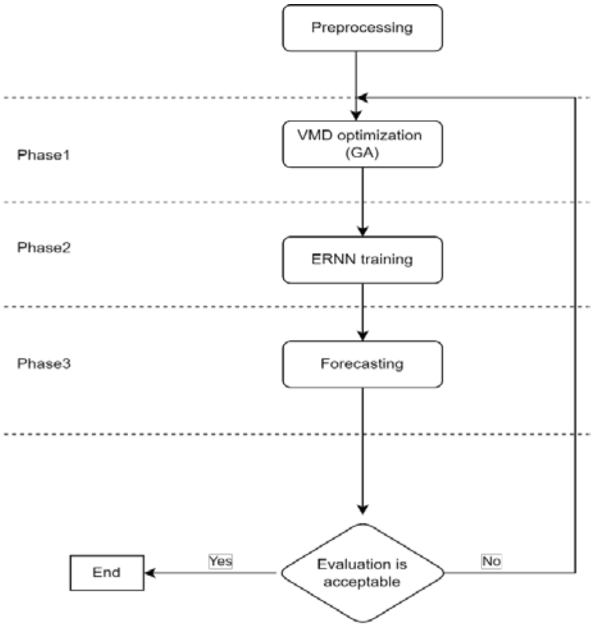


Fig. 2. A General Methodology Flowchart in TEC Forecasting.

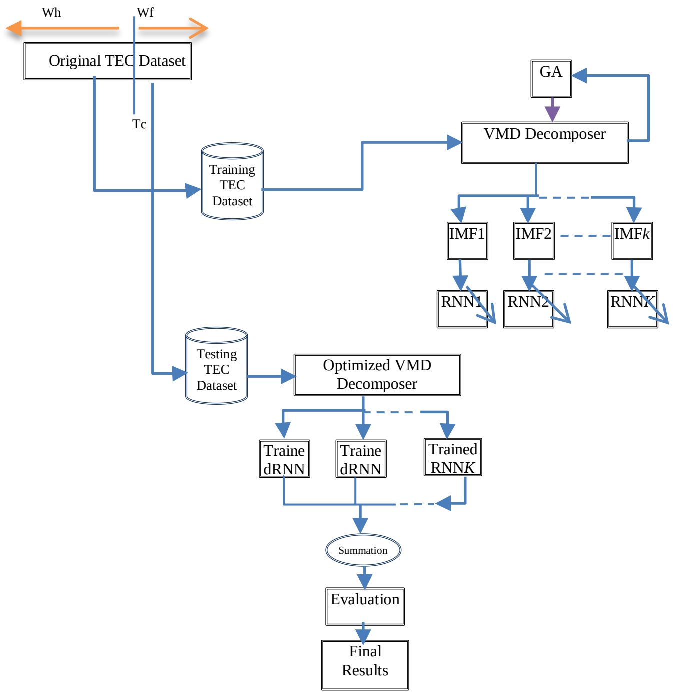


Fig. 3. Training and Testing Phases of TEC Forecasting.

E. Data Description

Data are recorded by the GPS Ionosphere Scintillation and TEC Monitor (GISTM) with a dual-frequency receiver GSV4004B at UKM station, geographic coordinate: 2.55 °N, 101.46 °E. At the L1 (1575.42 MHz) and L2 (1227.6 MHz) frequency bands, the GSV4004B receiver can track up to 11 GPS satellites. Amplitude and phase are monitored at 50 Hz, while code/carrier divergence (C/No) is sampled at 1 Hz for each satellite. The GPS Ionospheric Scintillation and TEC Monitor (GISTM) shows ionospheric delay over Universiti Kebangsaan Malaysia (UKM) station from 2011 to 2013.

F. Parameter Setting

GA optimization has been employed to find the best values of K and α. Considering that the optimization process consumes considerable computational time, only a few numbers of individuals were permitted to participate. For GA, the number of iterations is set to 6, the number of individuals to 5, the searching range for K between 2 and 16, and the searching range for α between 1,000 and 10,000 for GA optimization. The parameters are shown in Table I.

The proposed method has also been compared to ARIMA using the parameters mentioned in Table II. The observation's lag time is set to 30, the degree of difference is set to 2, and the moving window size is set to 15.

TABLE I. PARAMETER SETTING FOR GA

Dataset	No. of iterations	Crossover prob.	Mutation prob.	***K***	***α***
2011-2012	6	0.5	0.3	[2,16]	[1000,10000]
2013	10	0.5	0.3	[2,16]	[1000,10000]

TABLE II. ARIMA PARAMETER SETTINGS

Parameter name	**Value**
Number of lag observations (p)	30
Degree of differencing (d)	2
Moving average window size	15

For ENN, the number of epochs was set to 6 and 10 for 2011 & 2012 and 2013, respectively. The number of neurons in the first layer was set to 60; the second layer was set to 30. The time window lag was set to 30. The parameter settings are shown in Table III.

TABLE III. ELMAN RNN PARAMETER SETTINGS

Parameter name	**2011 &2012**	**2013**
Number of epochs	6	10
Neurons present in the First layer	50	50
Neurons in the hidden layer	30	25
Time-window (time lag)	30	30

G. Performance Metrics

The evaluation is based on two measures:

The first one is the mean squared error (MSE) which is given to evaluate the prediction of one time series, and is given by Equation 7.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2 \qquad (7)$$

where, Y_t denotes the ground truth value of the time series at moment t, and F_t denotes the predicted value of the time series.

The second metric is the improvement percentage from one model to another, and it is used to evaluate the relative improvement of proposed model over the benchmark as shown in Equation 8.

$$Percentage = \left|\frac{MSE_{Our} - MSE_{RMSE}}{MSE_{RMSE}}\right| \qquad (8)$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

This section presents the obtained results from the experiments conducted in this study and analysis the results of the developed OVMD-RNN and its comparison with the benchmark ARIMA.

For the 2011 and 2012 datasets, 21 months were used for training and the remaining three months for testing. Fig. 4 depicts the original TEC time series for 2011 and 2012 before applying the proposed model, and the non-stationarity (the frequency varies over time) can be seen. Fig. 5 shows the testing data's TEC time series for three consecutive months.

For further evaluation, the proposed model has been tested on another dataset for the year 2013. Fig. 6 depicts the original TEC time series for 2013 before applying the proposed model. Fig. 7 is the TEC time series for two consecutive months for the testing data.

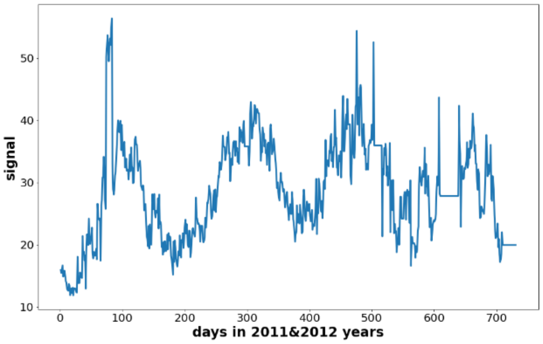


Fig. 4. TEC for 2011 and 2012 Dataset.

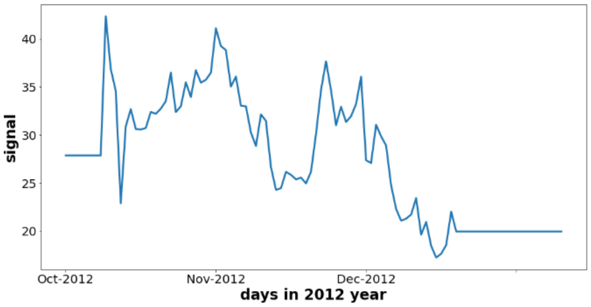


Fig. 5. TEC Time Series for Three Consecutive Months for Testing Data.

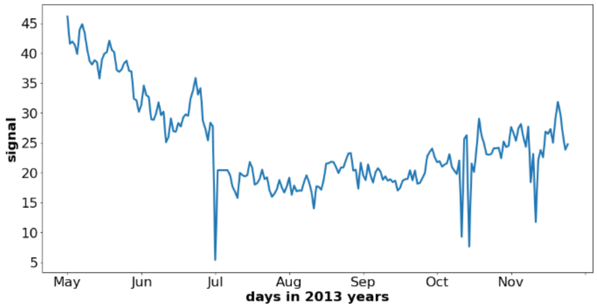


Fig. 6. TEC for 2013 Dataset.

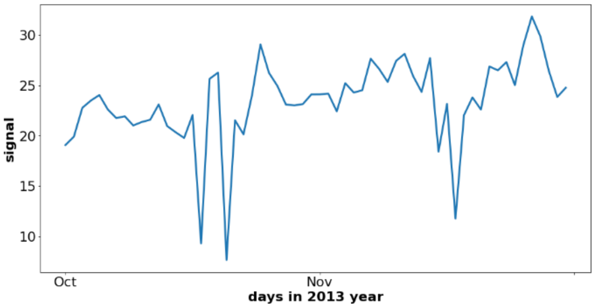


Fig. 7. TEC Time Series for Two Consecutive Months for Testing Data.

A. VMD Decomposition

1) K=5: Firstly, the TEC time series for 2011-2012 has been decomposed into its components using VMD and analyzed into five IMFs, which representing the optimal number of K components obtained through GA optimization, as shown in Table IV. Fig. 8 shows K=5 components for the test TEC time series data part.

2) K=7: As shown in Fig. 9, the testing TEC time series for 2013 was decomposed into its components and analysed into seven IMFs representing the optimal number of K components obtained through GA optimization.

B. Forecasting

Fig. 10, for example, displayed the visualisation of mode 2 forecasting and their comparison to real values. The visualization of mode 4 forecasting and its comparison to real values are depicted in Fig. 11. The visualisation shows that the proposed model OVMD-RNN can follow all the up and down peaks in the original data.

TABLE IV. THE BEST K AND α VALUE BY THE GA

Best *K*	**Best α**	**Algorithm**
5	9948	GA

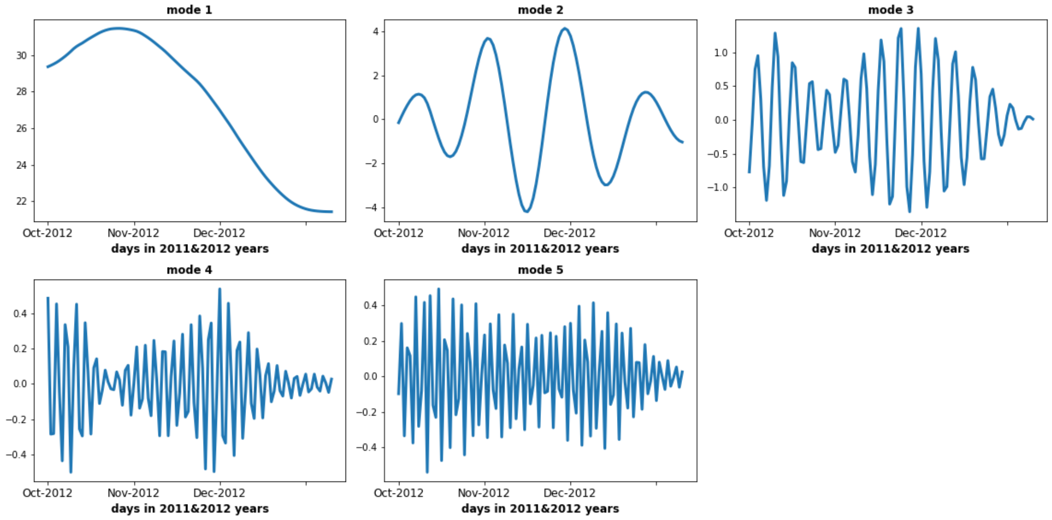


Fig. 8. VMD Decomposition for K=5 for Testing Data.

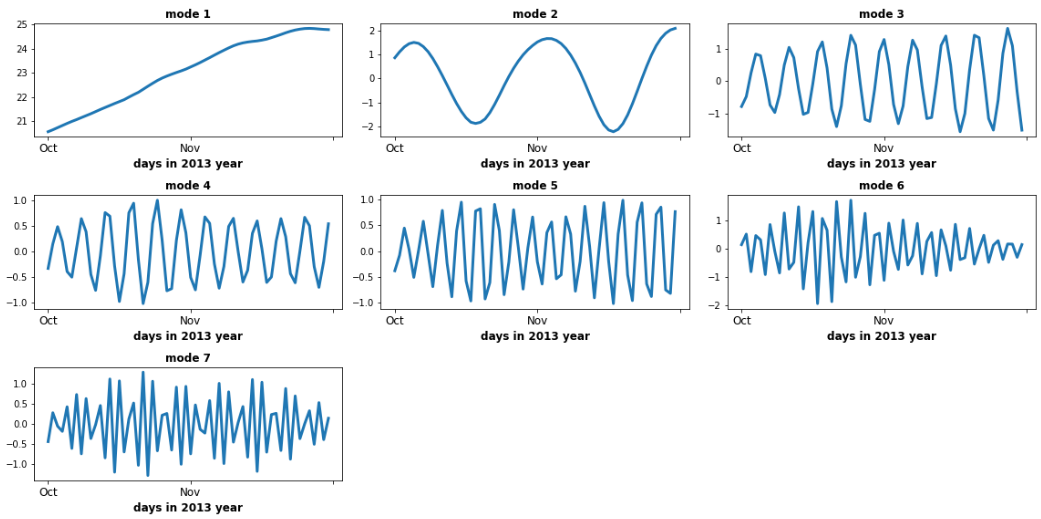


Fig. 9. VMD Decomposition for K=7 for Testing Data.

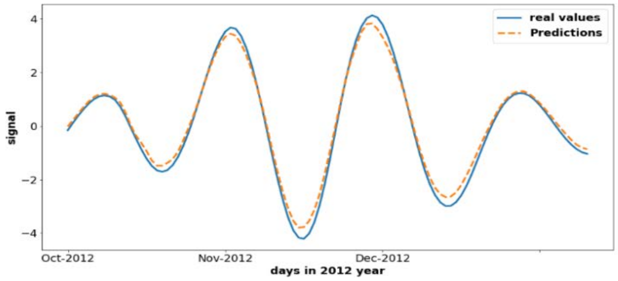


Fig. 10. Real Values Vs Prediction of VMD Mode 2.

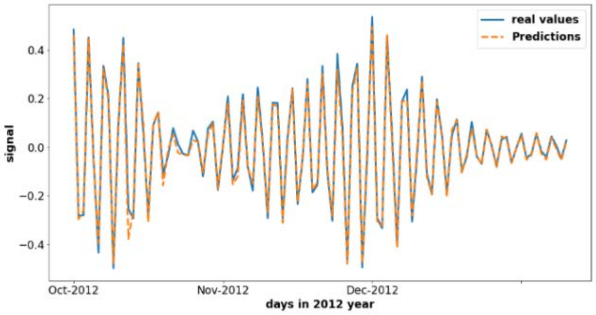


Fig. 11. Real Values Vs Prediction of VMD Mode 4.

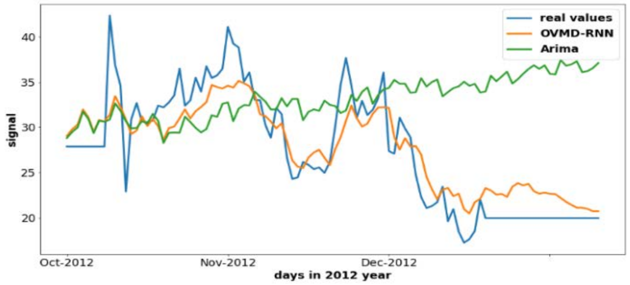


Fig. 12. A Comparison between Real Values, OVMD_RNN and ARIMA for K = 5.

The forecasting results have been presented in Fig. 12. It is observed from the figure that in the three models, the proposed model has provided better forecasting compared with ARIMA, which has shown a lack of capturing the trend and the pattern in the original time series. Furthermore, it has been discovered that the proposed model correctly predicted the four peaks in the original data, which show a level of 40 in October and November and a little lower level in late November and early December.

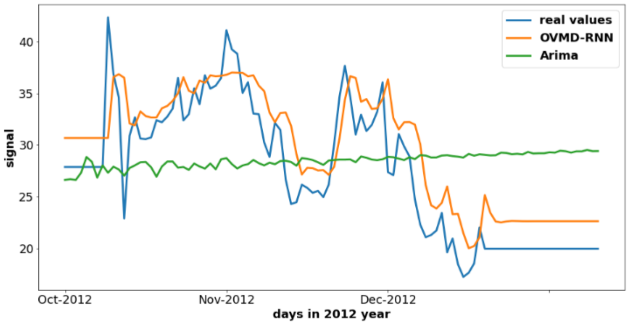


Fig. 13. A Comparison between Real Values, OVMD_RNN and ARIMA for K = 1 (without VMD).

Fig. 13 presented the forecasting results for the years 2011&2012 in the last three months. Without applying VMD (K=1), observing Fig. 12, 13, it is noticed that K=1 has better tracking than a higher loss at the training. VMD interprets it has caused the removal of an essential part of the signal, which has led to missing by the neural network while training for K= 5, when K=1, this has not been observed because this part was preserved when training by ENN. This provides that VMD based training might not always provide good results due to the residual signal that is deleted by this process.

For further evaluation, the forecasting behavior in the year 2013 is presented for the proposed model and its comparison with ARIMA.

As is shown in Fig. 14 and 15 increasing the value of K has enabled better prediction. Furthermore, the model has successfully tracked all the up and down patterns in the time series, ranging between 15 and 30.

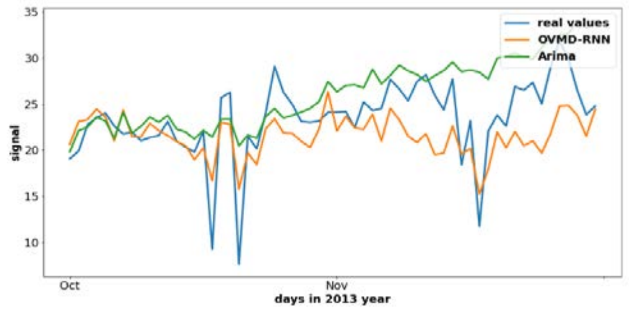


Fig. 14. A Comparison between Real Values, OVMD_RNN and ARIMA for K = 7.

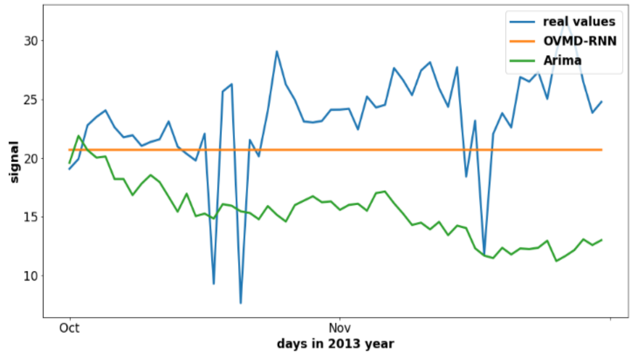


Fig. 15. A Comparison between 2013 Real Values, OVMD_RNN and ARIMA for K = 1 (without VMD) and K = 7.

Fig. 14 presented the forecasting results for the year 2013 in the last two months because the first months have been used for training. It is observed from the figure that the proposed model has provided better forecasting compared with ARMIA, which has shown a lack of capturing the trend and the pattern in the original time series. In addition, it is noticed that the proposed model has successfully forecasted all the up and down peaks in the original data, which appears with the level of 10 for October and at the end of November with the level of 35.

Fig. 15 presented the forecasting results for the year 2013 in the last two months. Without applying VMD (K=1), the model can't track the trend of the time series.

C. Evaluation

The evaluation was focused on creating the loss value and comparing it to the ARIMA benchmark.

1) Loss value: The loss value in the training phase is presented in Fig. 16, which shows that K=1 has the highest loss value than K=5, which is the value resulting from GA. In addition, for leading the optimality of GA.

Observing Fig. 13, it is noticed that K=1 has better tracking than a higher loss at the training. In K=5, the VMD process removed some important parts and caused some missing during Elman training. The result shows VMD-based training might not always provide good results due to the residual signal deleted by the process.

The loss value in the training phase is presented in Fig. 17 for the dataset 2011-2012, which proves the superiority of the proposed forecasting model when the value of K is selected optimally.

Regarding the loss value K=5, the proposed model has been compared to ARIMA, as shown in Fig. 17. The proposed approach has a smaller error or loss value than ARIMA. The results demonstrate that it is superior.

To summarize the performance, the overall loss value K =1 has been presented, which indicates VMD decomposition and K= 7, which suggests the result of GA.

In Fig. 14 and 15, it is observed that increasing the value of K from 1 to 7 has enabled a lower value of loss than ARIMA, which proves the superiority of the proposed forecasting model when the value of K is selected optimally.

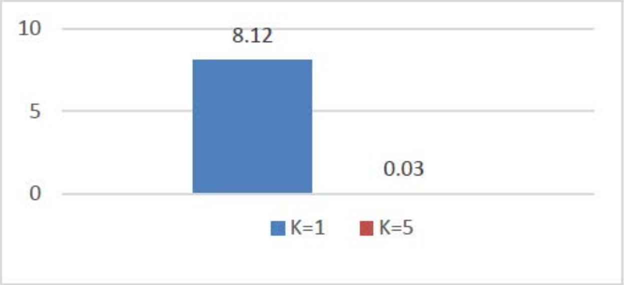


Fig. 16. The Loss Value of OVMD-RNN for When K =1, K=5.

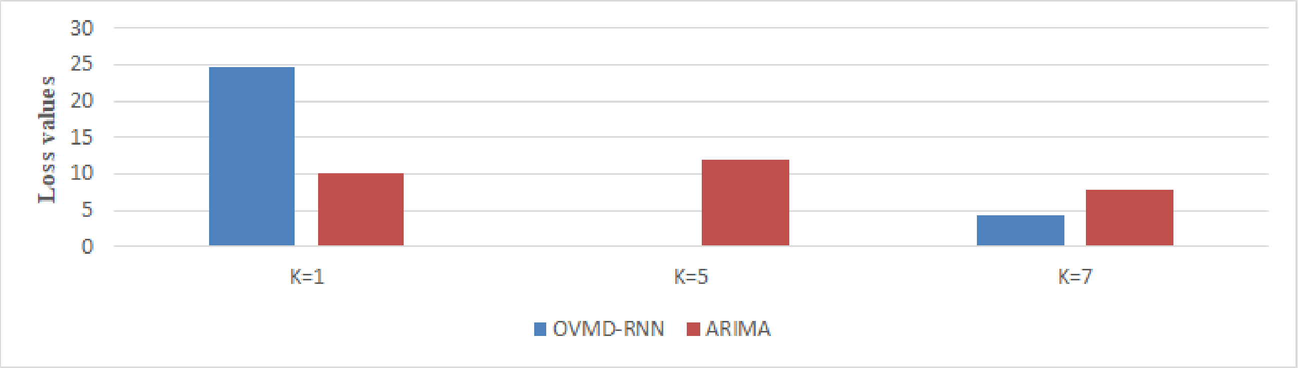


Fig. 17. The Loss Value of OVMD-RNN Vs ARIMA.

2) Accuracy: In terms of the loss value K = 5, OVMD-RNN has been compared to ARIMA, as shown in Fig. 12, Because proposed approach has a smaller error or loss value than ARIMA, the results demonstrate that it is superior. The improvement percentage is 99% as calculated based on Equation 9.

$$Percentage = \left|\frac{RNN-ARIMA}{ARIMA}\right| = \left|\frac{0.05-12.47}{12.47}\right| = 0.99 \qquad (1)$$

TABLE V. OVERALL PREDICTION COST FOR ARIMA AND OVMD-RNN BASED ON DIFFERENT YEARS AND VALUES OF K

	2011&2012		2013	
K	1	5	1	7
ARIMA	6.92	12.04	10.14	7.79
OVMD-RNN	8.12	0.03	24.72	4.29
Improvement percentage	0.17	99%	%143	%44

From Table V, K = 1 indicates that VMD has not been applied. So, the non-stationary time series were not divided into stationary components but were predicted directly using ENN and ARIMA. For K = 5 and K = 7, these values were obtained by implementing GA on both times series 2011-2012 and 2013, respectively, to find the best K value. These K values provide the lowest loss value.

D. Observations

From the results obtained, it has been observed that TEC is a non-stationary time series making it challenging to forecast. Moreover, VMD is a good candidate for decomposing TEC time series into stationary components. Still, sometimes VMD-based training might not always provide good results due to the residual signal.

The proposed model (OVMD-RNN) does not implement multi-time series collected from different areas to be generalized. Therefore, extending the model to accept multi-time series at one time will enable more accurate forecasting; this can be applied for future work.

V. SUMMARY AND CONCLUSION

This paper has created a novel forecasting approach for the TEC time series. VMD is used to split the original TEC time series into necessary stationary components, considering the non-stationarity of the data and the need to include non-linear knowledge for forecasting. Each essential TEC component was trained and forecasted using an Elman RNN. In addition, the method consists of an optimization algorithm for determining the best VMD decomposer parameters. The VMD parameters, K and α selection, utilized the GA optimization method. The GPS Ionospheric Scintillation and TEC Monitor (GISTM) with a dual-frequency receiver GSV4004B at the UKM station evaluated our obtained dataset for three years, 2011, 2012 and 2013. The evaluation was focused on creating the loss value and comparing it to the ARIMA benchmark. It showed that the proposed progressive technique with two decomposition values for K = 4 and 5 and a significant reduction of the loss value was superior. Future research will develop multi-dimensional TEC forecasting from multiple places within the same geographic region.

ACKNOWLEDGMENT

We like to thank the Data Mining and Optimization (DMO) Lab of Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), College of Science for Women, University of Babylon, Iraq for the expert's knowledge sharing, and the Space Science Centre (ANGKASA) UKM, for providing the data. Massive appreciation to UKM for providing grant GGP-2020-032 for funding this research and the Iraq Government for sponsoring the postgraduate study.

REFERENCES

[1] L. Mallika, D. V. Ratnam, S. Raman, and G. Sivavaraprasad, "Machine learning algorithm to forecast ionospheric time delays using Global Navigation satellite system observations," Acta Astronautica, vol. 173, pp. 221-231, 2020.

[2] U. Sezen, T. L. Gulyaeva, and F. Arikan, "Online international reference ionosphere extended to plasmasphere (IRI-Plas) model," in 2017 XXXIInd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS), 2017, pp. 1-4: IEEE.

[3] A. A. Akyol, O. Arikan, and F. Arikan, "A machine learning-based detection of earthquake precursors using ionospheric data," Radio Science, vol. 55, no. 11, pp. 1-21, 2020.

[4] D. V. Ratnam, Y. Otsuka, G. Sivavaraprasad, and J. K. J. A. i. S. R. Dabbakuti, "Development of multivariate ionospheric TEC forecasting algorithm using linear time series model and ARMA over low-latitude GNSS station," vol. 63, no. 9, pp. 2848-2856, 2019.

[5] M. Kaselimi, A. Voulodimos, N. Doulamis, A. Doulamis, and D. J. R. S. Delikaraoglou, "A Causal Long Short-Term Memory Sequence to Sequence Model for TEC Prediction Using GNSS Observations," vol. 12, no. 9, p. 1354, 2020.

[6] S. Inyurt, M. Hasanpour Kashani, and A. Sekertekin, "Ionospheric TEC forecasting using Gaussian Process Regression (GPR) and Multiple Linear Regression (MLR) in Turkey," Astrophysics and Space Science, vol. 365, no. 6, p. 99, 2020/06/10 2020.

[7] K. Ansari, S. K. Panda, O. F. Althuwaynee, and O. Corumluoglu, "Ionospheric TEC from the Turkish Permanent GNSS Network (TPGN) and comparison with ARMA and IRI models," Astrophysics Space Science, vol. 362, no. 9, pp. 1-14, 2017.

[8] S. Salma, G. Sivavaraprasad, B. Madhav, and D. V. Ratnam, "Implementation of VARMA Model for Ionospheric TEC Forecast over an Indian GNSS Station," in 2020 5th International Conference on Devices, Circuits and Systems (ICDCS), 2020, pp. 144-148: IEEE.

[9] L. R. Cander, "Ionospheric space weather forecasting and modelling," in Ionospheric space weather: Springer, 2019, pp. 135-178.

[10] J. B. Habarulema, L.-A. McKinnell, and P. J. Cilliers, "Prediction of global positioning system total electron content using neural networks over South Africa," Journal of Atmospheric solar-terrestrial physics vol. 69, no. 15, pp. 1842-1850, 2007.

[11] J. B. Habarulema, L.-A. McKinnell, P. J. Cilliers, and B. D. Opperman, "Application of neural networks to South African GPS TEC modelling," Advances in Space Research, vol. 43, no. 11, pp. 1711-1720, 2009.

[12] G. Sivavaraprasad and D. V. Ratnam, "Performance evaluation of ionospheric time delay forecasting models using GPS observations at a low-latitude station," Advances in Space Research, vol. 60, no. 2, pp. 475-490, 2017.

[13] N. Elmunim, M. Abdullah, A. Hasbi, and S. Bahari, "Comparison of GPS TEC variations with Holt-Winter method and IRI-2012 over Langkawi, Malaysia," Advances in Space Research, vol. 60, no. 2, pp. 276-285, 2017.

[14] G. Sivavaraprasad, D. V. Ratnam, M. Sridhar, and K. Sivakrishna, "Modelling and forecasting of ionospheric TEC irregularities over a low latitude GNSS station," Astrophysics Space Science, vol. 365, no. 10, pp. 1-14, 2020.

[15] Z. Huang, Q. Li, and H. Yuan, "Forecasting of ionospheric vertical TEC 1-h ahead using a genetic algorithm and neural network," Advances in Space Research, vol. 55, no. 7, pp. 1775-1783, 2015.

[16] M. R. G. Razin and B. Voosoghi, "Modeling of ionosphere time series using wavelet neural networks (case study: NW of Iran)," Advances in Space Research, vol. 58, no. 1, pp. 74-83, 2016.

[17] J. R. K. K. Dabbakuti, A. Jacob, V. R. Veeravalli, and R. K. Kallakunta, "Implementation of IoT analytics ionospheric forecasting system based on machine learning and ThingSpeak," IET Radar, Sonar Navigation, vol. 14, no. 2, pp. 341-347, 2019.

[18] G. Li, X. Ma, and H. Yang, "A hybrid model for monthly precipitation time series forecasting based on variational mode decomposition with extreme learning machine," Information, vol. 9, no. 7, p. 177, 2018.

[19] R. Jegadeeshwaran, V. Sugumaran, and K. Soman, "Vibration based fault diagnosis of a hydraulic brake system using Variational Mode Decomposition (VMD)," Structural Durability Health Monitoring, vol. 10, no. 1, p. 81, 2014.

[20] C. Zhao et al., "Novel method based on variational mode decomposition and a random discriminative projection extreme learning machine for multiple power quality disturbance recognition," IEEE Transactions on Industrial Informatics, vol. 15, no. 5, pp. 2915-2926, 2018.

[21] M. Niu, Y. Hu, S. Sun, and Y. Liu, "A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting," Applied Mathematical Modelling, vol. 57, pp. 163-178, 2018/05/01/ 2018.

[22] M. A. A. Albadr, S. Tiun, M. Ayob, F. T. Al-Dhief, K. Omar, and F. A. Hamzah, "Optimised genetic algorithm-extreme learning machine approach for automatic COVID-19 detection," PloS one, vol. 15, no. 12, p. e0242899, 2020.

[23] Mirjalili, S., Song Dong, J., Sadiq, A.S. & Faris, H. 2020. Genetic algorithm: Theory, literature review, and application in image reconstruction. Nature-Inspired Optimizers 69–85.

[24] Kamanditya, B. & Kusumoputro, B. 2020. Elman recurrent neural networks based direct inverse control for quadrotor attitude and altitude control. In 2020 International Conference on Intelligent Engineering and Management (ICIEM). pp. 39–43. IEEE.

[25] G. Ren, Y. Cao, S. Wen, T. Huang, and Z. Zeng, "A modified Elman neural network with a new learning rate scheme," Neurocomputing, vol. 286, pp. 11-18, 2018.

[26] R. Hannah Jessie Rani and T. Aruldoss Albert Victoire, "A hybrid Elman recurrent neural network, group search optimization, and refined VMD-based framework for multi-step ahead electricity price forecasting," Soft Computing, vol. 23, no. 18, pp. 8413-8434, 2019.

Assaying the Statistics of Crime Against Women in India using Provenance and Machine Learning Models

Geetika Bhardwaj, Dr. R. K. Bawa
Department of Computer Applications, Punjabi University, Patiala, India

Abstract—**Now-a-days, the surging of crime against women is occurring at a startling rate in India. According to the National Commission for Women, there was a 46% increase in reports of crimes against women in the initial months of the year 2021 in comparison with the same period in 2020. However, to handle this problem, the need of the hour is to fetch relevant and timely information about the various types of crime taking place and make specific predictions based on the existing information to safeguard women from future predictable contingencies. AI and Machine learning mechanisms have become a powerful tool in predicting the crime rate in India under various crime categories by analyzing the crime patterns, crime–centric areas, and the comparative study of various crime categories. Hence, from 2001 to 2019, a women's crime-based dataset from NCRB has been used in this paper, which included various crime sub-categories, for instance; molestation, sexual harassment, rape, kidnapping, dowry deaths, cruelty to family, importation of girls, immortal traffic, sati prevention act, and others. To acquire a better understanding of the data, a framework has been created which makes use of provenance and machine learning algorithms on the dataset, which has been grouped based on several factors such as distribution of cases convicted or reported every year, safest and un-safest states for women in India, etc. Different machine learning algorithms, such as gradient boosting and its many versions, Random forest, and many more, have been used on the dataset. Their performances are evaluated using various metrics such as accuracy, recall, precision, F1 score, and root mean error square.**

Keywords—Crime against women; provenance; scalar techniques; machine learning techniques; decision tree; random forest; gradient boosting; XgBoost; CatBoost; LightGBM

I. Introduction

The overall crime rate in India is increasing at a steady pace. Crime cannot be predicted as it is either efficient, coincidental, or goes unreported. Various modern advancements and hi-tech procedures enable criminals to carry out their crimes when it comes to cybercrime. As far as women's crimes are concerned, it has been seen that a girl can be a victim or target of a crime from the moment she is born or even before. According to crime statistics, authority violations against women, for example, chain grabbing, sex abuse, child abuse, assault, and murder, are rapidly increasing [1]. Hence, keeping that in view, the establishment of the gender equality principle was done in the Constitution of India, and to uphold and implement the Constitutional mandate, the state has created several laws as well as has taken various actions to guarantee equal rights, eradicate social injustice, and prohibit multiple forms of violence and massacres [2].

India's National Crime Records Bureau records various incidents showing that crime against women increased by 6.4%, and it occurs every three minutes [3]. The reports also revealed that in 2011 the number of reported crimes against females was more than 228,650. In the year 2015, reported incidents were more than 300,000. It can be deduced that there was a rise of 44% in felonies against the members of fair sex.7.5% of females residing in the state of West Bengal in India account for 12.7% of reported crimes against women [4]. The female population of Andhra Pradesh accounts for 7.3% of total India's women population, and 11.5% of all reported crimes against women were from this state. Hence to compile all the information, a graphical analysis has been shown in Fig. 1 to define the rate of women crimes in India in different states for 2018. It is also important to note that exact figures on the breadth of case occurrences are difficult to obtain because many cases go unreported. This is mainly due to the potential reporter's fear of scorn or embarrassment and tremendous pressure not to jeopardize the family's honor [5].

A crime investigation system should be able to rapidly and efficiently identify crime patterns to detect and act on future crime trends to work on such crimes. Various law enforcement agencies and state governments should take significant steps to reduce such crimes and promote a secure environment for women [6]. Multiple scholars have been working on detecting women-related crime worldwide in the field of research. Data Analytics and machine learning have contributed notably to detecting and preventing crime, providing a basis for crime analysis. They are also acknowledged as a relatively new and highly sought-after area of research. In reality, law enforcement agencies are seeking the support of data mining and AI techniques to help deter crime and ensure law enforcement. In a nutshell, data mining is a branch of the multidisciplinary subject of database knowledge discovery. The input in the process of Data mining is the unprocessed data, which is transformed into information that is used to produce precise projections and applied to real-life circumstances (through inference and analysis). Multiple techniques, such as scientific and statistical machine learning algorithms, have recently been used in the recognition of images and speech, detection of medical ailments, and classification [7] and are additionally designed to estimate crime rates for a particular year formed on statistical information of crime against females.

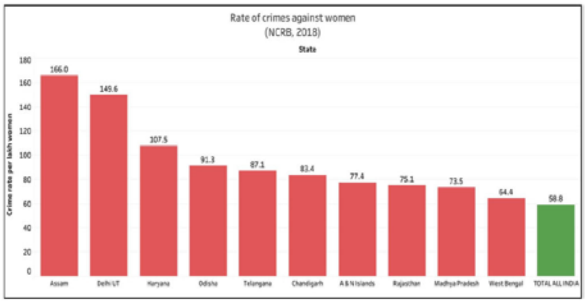


Fig. 1. Rate of Crimes against Women in the Year 2018.

In a nutshell, it can be said that studying the crime data helps us to solve criminal cases and is taken as a first step to prevent the1 crime via which we can reduce or deter criminals and their activities. But depending on raw data is utterly erroneous as it contains plenty of noise, incomplete or missing values, etc. to confuse the investigation team so that they cannot afford either forecast its future features or catch the culprit. Hence, to work on such destructed data, we want to create a decision-making system that can analyze the crime data, find the data uncertainty, and remove the errors.

The essential part and main motive behind this research work is data provenance which is a novelty and is used to train the system so that we can work on the uncertainty of the crime data against women. Provenance aids in getting into the origin of data and various transformations made on the data over time. Suppose the analysis is done based on data obtained from a particular source that no longer exists or is unavailable to us; in that case, the data provenance generated might be the guiding light to establish the uncertainty in the data.

Hence, the contribution of the research for analyzing the statistics of women-based crimes in India using crime data provenance is as follows:

1) The information is gathered from NCRB and Kaggle between 2001 to 2019, which is further pre-processed to remove NAN or missing values.

2) Later feature scaling techniques such as Min-Max, Standard scalar, and PCA is being applied to scale or generalize the values.

3) At the end, the scaled data is further evaluated using various evaluation parameters such as accuracy, rmse, R2, mse, precision, recall, etc to find the best technique for predicting the crime.

The study has been divided into several sections where Section I has been already defined as an Introduction; Section II defines the related work. The proposed model is demonstrated in Section III, which includes information about datasets, libraries, data collection, data provenance, data pre-processing, feature scaling, algorithms, and evaluative parameters. Section IV mentions the results, whereas Section V concludes the study.

II. Related Work

Many scholars have investigated crime control concerns and proposed various crime prediction algorithms. The qualities of the techniques and the dataset, along with the challenges, have been used as a benchmark (Table I) to determine the accuracy of the prediction and generate the research gap. In [7], the authors examined and found the key factors influencing crime in certain parts of the country. They employed clustering, a graphical representation based on determining which areas have the greatest and minor crimes. In addition, the authors have implemented the Community Detection Algorithm. The authors presented a method for predicting crime without human involvement [8] using computer vision and machine learning technologies. The paper employed rectified linear units (ReLU) and convolutional neural networks (CNN) to identify weapons in images, such as knives or guns. This aided in confirming the occurrence of a crime and identifying the event's site. The results looked highly accurate, with roughly 92% accuracy for a test set. In [9], the authors developed a provenance capturing mechanism that aims to trace digital evidence's transformation to bolster the trustworthiness of digital evidence when the incident response takes place on the affected system. In their work, the data provenance was recorded simultaneously in both the systems, i.e., the incident response system and the affected system.

Likewise, in [10], the authors employed a variety of machine learning algorithms on data of criminal cases in India to find patterns in criminal activities in a particular geographical location. The aim was to reduce the number of pending criminal cases by classifying them based on their crime patterns to solve them faster. Different machine learning algorithms were applied, and a comparison was made based on several parameters to find the best algorithm to solve this problem efficiently. It had been concluded that the Random Forest Classification method was best suited to predict the desired results after classification. In [11], the authors created a prediction model that can be used for the prognosis of crime rates accurately, and they tested the accuracy of six different types of Machine learning algorithms on crime data, which included Linear Regression, SVM, KNN, decision trees, Nave Bayes and CART (Classification and Regression Tree). The authors of [12] described how the frequency of crimes and crime features in India, such as rape, sexual assault, and kidnapping, may be examined using machine learning models in the investigation process.

Similarly, in [13], the authors worked on several machine learning algorithms predicting crimes. They devised a new framework to combine two machine learning algorithms, i.e., XGBoost and TF-IDF, to improve the results of various algorithms used in text mining to predict crimes. The improvement of data accuracy to strengthen the accuracy of crime prediction overweighs the optimization process of algorithms. The inaccuracy of crime data might result in an inaccurate prediction of a specific category of crime that has its basis in the historical data. Training a good classification model is imperative in improving the accuracy of data, which in turn forms the basis for analyzing and accurately predicting crime. The text-based classification of theft crime data based on the two algorithms, i.e., XgBoost and TF-IDF, were also used to get a logical and error-free classification effect of data. This was a constructive trial at a machine learning algorithm for data mining of police-related data and quintessential for predicting crime.

TABLE I. ANALYSIS OF THE PREVIOUS WORK

Ref	Dataset	Techniques	Outcomes	Limitation
[8]	Dataset containing criminal tools	CNN, ReLu	Accuracy = 90.02%	High computational cost
[11]	Real world data	KNN	Mean = 0.33	Accuracy needed to be improved
[34]	Crime reports	Visionary system, natural language processing	The system detected the crime based on analysing analytic provenance trails.	Multiple techniques needed to be incorporate for achieving the better results
[30]	NCRB data	Linear Regression	Acc = 83%	Limited dataset

[13]	data collected from 2009 to 2019	Tf-IDF, XGBoost model	Prec = 92.3%
			Rec = 91.6%
			F1 score = 91.9%
				Values of performance metrics needs to be enhanced

[31]	Data collected from past 10 years	Logistic Regression	Acc = 80.769%	Less scalability
		Random forest	Acc= 76.923%	
		Naïve Bayes	Acc = 80.769%	
		Decision tree	Acc= 76.923%	
[33]	Crime dataset	Huber Regression	Results obtained in the form of predictions and score of each state where women crime has taken place	Design complexity

No research on data provenance classification in criminal data has been conducted. Scientific data is kept in databases. Therefore, provenance management solutions were created with that in mind. No system has taken the provenance of crime data and its implementation concerns and challenges into account. Various women-based crimes have been targeted using multiple machines and deep learning techniques. Still, it has been discovered that researchers have encountered specific issues, either in terms of detection or system performance. A few models, such as logistic regression, KNN, Relu, and SVM, had a complex design, used a limited dataset, or needed improvement in their performance. As a result, the research's primary motivation is to fill the gaps to forecast a better system for detecting crime utilizing data provenance.

III. PROPOSED SYSTEM

A framework has been designed to analyze the statistics of women-based crimes in India, named the Crime Data Provenance Framework, as shown in Fig. 2. This framework aims to design a methodology to use the provenance of the given crime data in the form of annotations (a subtext or metadata which provides additional information about the actual data) to create crime classes that are used further for the classification and prediction of crime data. There are a variety of crime sub-categories in this framework which has become the basic building block, and prediction models have been applied. The results have been compared and analyzed. The framework has only been implemented for a single category of crime, i.e., crime against women. The various steps in the framework are categorized into several sub-sections wherein each sub-section elucidates the purpose and tasks performed in each step.

A. Data Collection

The data has been accumulated from the government website named National Crime Record Bureau (NCRB) [35] and Kaggle [36]. The website was established on 11th March 1986. The general aim was to empower Indian Police with Information Technology and Criminal Intelligence to make law enforcement more productive. The data from 2001 to 2015 was in .xls format, and the rest from 2015 to 2019 was in pdf format. Some books helped to understand how crime provenance trends can be represented in a textual form that could later become part of Provenance. The first book, named 'Violence Against Women in India' prepared by International Center for Research on Women, aimed at elaborating on the trends and patterns in crime against women, and the second book was 'Tackling Violence Against Women: A Study of State Intervention Measures' which was a report prepared by Bhartiya StreeShakti, a voluntary, autonomous, apolitical organization committed to empowering women, families and the society at large. The report aimed to study different aspects of crime against women by taking perspectives from other professionals like Lawyers, Police Personnel, Public Prosecutors, and Medical Officials [12].

At a fleeting glance, the data seemed to have a lot of dimensions as there were a lot of crime subcategories. There were several crime categories as shown in Fig. 3 and each category had further subcategories. Each crime subcategory had different crime status categories. For instance, our crime category was CRIME AGAINST WOMEN. Further, it had subcategories such as cruelty by husband and relatives, kidnapping of women and girls, Indecent Representation of Women (Prohibition) Act, 1986, and many more.

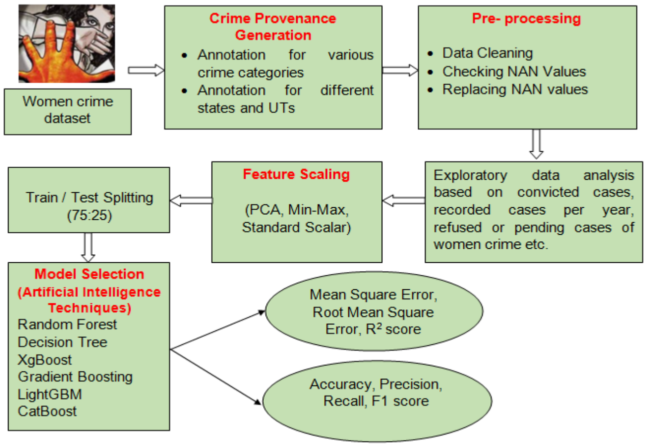

Women crime dataset
Crime Provenance Generation
• Annotation for various crime categories
• Annotation for different states and UTs
Pre- processing
• Data Cleaning
• Checking NAN Values
• Replacing NAN values
Exploratory data analysis based on convicted cases, recorded cases per year, refused or pending cases of women crime etc.
Feature Scaling
(PCA, Min-Max, Standard Scalar)
Train / Test Splitting (75:25)
Model Selection (Artificial Intelligence Techniques)
Random Forest
Decision Tree
XgBoost
Gradient Boosting
LightGBM
CatBoost
Mean Square Error, Root Mean Square Error, R² score
Accuracy, Precision, Recall, F1 score

Fig. 2. Crime Data Provenance Framework.

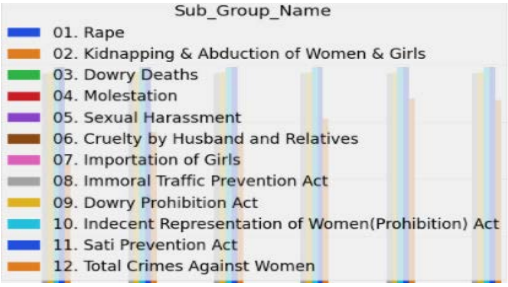


Fig. 3. Various Crime Categories under Crime against Women.

B. Provenance Generation

The crime against women has been subcategorized based on physical or mental abuse, or both caused to women under various circumstances with varying intentions. Different laws are formulated to deal with cases for each crime category, and they have been listed along with their annotations as a small sample data in Table II.

In addition to this, the number of cases for different categories of crime reported at various places has also been organized and segregated according to various States and Union territories. As we know, India has 28 States and 8 Union Territories, and the data has been compiled by the number of cases of different crime categories reported in these places. The states and union territories have been annotated to be used as Provenance for crime prediction purposes. Table III consists of the name of a state and a Union Territory and their annotations as a sample.

TABLE II. SAMPLE OF ANNOTATIONS FOR DIFFERENT CRIME CATEGORIES

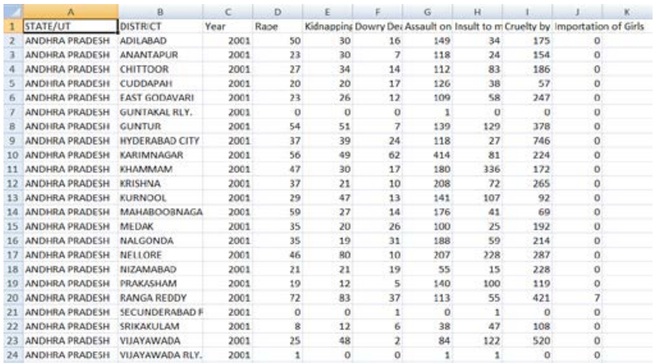
Case Sub-categories	Crime Act/Section	Case Code (Annotations)
Rape	Section 376 IPC	CSR-376
Kidnapping & Abduction of Women & Girls	Sec. 363-373 IPC	CSK-363-373
Commission of Sati Prevention Act 1987	NA	CSSP-1987
Protection of Women from Domestic Violence Act 2005	NA	CSDV-2005

TABLE III. SAMPLE OF ANNOTATIONS FOR DIFFERENT STATES AND UNION TERRITORIES

Name of the State/Union Territory	State/UT Code(Annotations)
Arunachal Pradesh	SCAR
Puducherry	UTPU
Himachal Pradesh	SCHP
Jharkhand	SCJH
Karnataka	SCKA
Kerala	SCKE
Madhya Pradesh	SCMP
Maharashtra	SCMH
Manipur	SCMA
Meghalaya	SCME
Mizoram	SCMI

C. Crime Data Pre-processing Phase

The Preprocessing of data is necessary to clean it and make it suitable for a machine learning model, which improves the effectiveness and precision of the machine learning model. Loading libraries and setting up the platform is the prerequisite to initializing the process of Data pre-Processing Fig. 4. Several Python libraries, such as Matplotlib, Numpy, Sklearn, Itertools, SimpleImputer, Seaborn, Maths, and Pandas, have been loaded to perform specific functions [14-16].



	A	B	C	D	E	F	G	H	I	J	K
1	STATE/UT	DISTRICT	Year	Rape	Kidnapping	Dowry De	Assault on	Insult to m	Cruelty by	Importation of Girls	
2	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	34	175	0	
3	ANDHRA PRADESH	ANANTAPUR	2001	23	30	7	118	24	154	0	
4	ANDHRA PRADESH	CHITTOOR	2001	27	34	14	112	83	186	0	
5	ANDHRA PRADESH	CUDDAPAH	2001	20	20	17	126	38	57	0	
6	ANDHRA PRADESH	EAST GODAVARI	2001	23	26	12	109	58	247	0	
7	ANDHRA PRADESH	GUNTAKAL RLY.	2001	0	0	0	1	0	0	0	
8	ANDHRA PRADESH	GUNTUR	2001	54	51	7	139	129	378	0	
9	ANDHRA PRADESH	HYDERABAD CITY	2001	37	39	24	118	27	746	0	
10	ANDHRA PRADESH	KARIMNAGAR	2001	56	49	62	414	81	224	0	
11	ANDHRA PRADESH	KHAMMAM	2001	47	30	17	180	336	172	0	
12	ANDHRA PRADESH	KRISHNA	2001	37	21	10	208	72	265	0	
13	ANDHRA PRADESH	KURNOOL	2001	29	47	13	141	107	92	0	
14	ANDHRA PRADESH	MAHABOOBNAGA	2001	59	27	14	176	41	69	0	
15	ANDHRA PRADESH	MEDAK	2001	35	20	26	100	25	192	0	
16	ANDHRA PRADESH	NALGONDA	2001	35	19	31	188	59	214	0	
17	ANDHRA PRADESH	NELLORE	2001	46	80	10	207	228	287	0	
18	ANDHRA PRADESH	NIZAMABAD	2001	21	21	19	55	15	228	0	
19	ANDHRA PRADESH	PRAKASHAM	2001	19	12	5	140	100	119	0	
20	ANDHRA PRADESH	RANGA REDDY	2001	72	83	37	113	55	421	7	
21	ANDHRA PRADESH	SECUNDERABAD F	2001	0	0	1	0	1	0	0	
22	ANDHRA PRADESH	SRIKAKULAM	2001	8	12	6	38	47	108	0	
23	ANDHRA PRADESH	VIJAYAWADA	2001	25	48	2	84	122	520	0	
24	ANDHRA PRADESH	VIJAYAWADA RLY.	2001	1	0	0	1	1	0	0	

Fig. 4. Preprocessing of Dataset.

D. Exploratory Data Analysis

The pre-processed data has been categorized into various categories, including the most unsafe and safe states for women in India from 2001 to 2013, as displayed in Fig. 5. The classification of different categories of women-based crimes, distribution of cases convicted per year, and distribution of cases reported per year (i.e., from 2001 to 2010) has been displayed in Fig. 6 and Fig. 7, which show that the convicted cases in terms of the total number of crimes against women range from 25000 to 35000 while as in the year 2010.

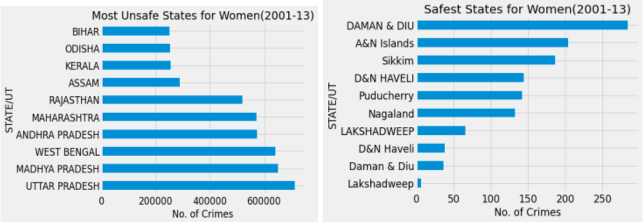


Fig. 5. Most Unsafe and Safe States for Women in India.

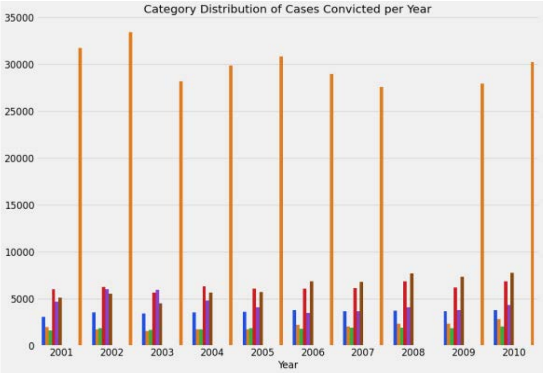


Fig. 6. Convicted Cases of Crime against Women in India.

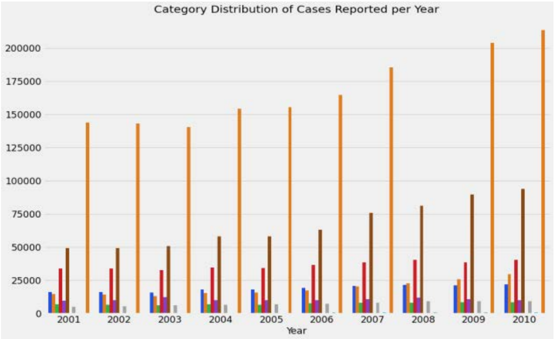


Fig. 7. Reported Cases of Crime against Women in India.

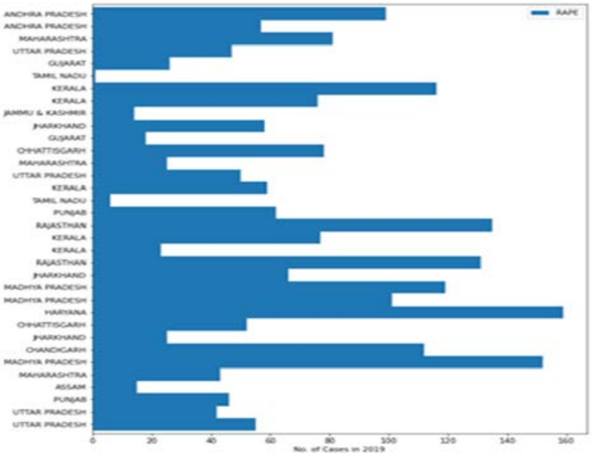


Fig. 8. Statewise Records of Rape Cases in 2019.

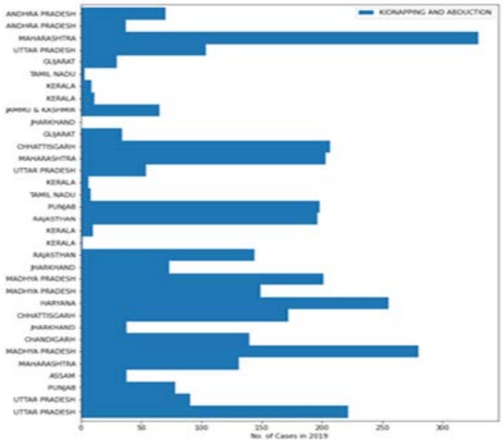


Fig. 9. Statewise Records of Kidnapping and Abduction in 2019.

In addition, Fig. 8 and Fig. 9 explore the data for the most recent year, 2019, the statistics displaying states that fell most heavily in crimes such as rape, the modesty of women, cruelty by husbands, kidnapping, and so on. The state of Haryana had the most rape cases, followed by Madhya Pradesh, while Tamil Nadu had the least. Similarly, Maharashtra ranked first in kidnapping and abduction, while Tamil Nadu and Kerala tied for last place. Dowry deaths have decreased dramatically across the country, with at least five states reporting the lowest number of women affected by dowries. Women's molestation has also reduced in several states, including Gujarat, Jammu & Kashmir, and many other conditions. Uttar Pradesh, Andhra Pradesh, and other states have many offenses involving spouse or relatives' maltreatment.

On the other hand, Andhra Pradesh ranks first among all states regarding girl importation as we know that the input to all these attributes is in numerical data. Hence, standardizing it using various scaling techniques mentioned in the next section is essential.

E. Feature Scaling

Feature scaling is the final stage in machine learning data processing. It is a method for standardizing the independent variables in a dataset within a given range. Multiple scaling techniques can be used here, but the one given priority is the one that offers more optimized results after normalizing the data [17-21]. Hence in this section, the scaling techniques such as Min-Max Scalar, Principal component analysis, and Standard Scalar have been used to showcase the performance of machine learning models such as decision tree, gradient boosting, and its many versions and random forest. These models have been applied to the dataset taken from various women-based crimes like cruelty by husbands, rape cases, an insult to modesty, kidnapping and abduction cases, dowry deaths, and importation of girls and are shown graphically in Fig. 10 to 12.

1) Min Max: The entire data is scaled between 0 and 1. To calculate min-max, the formula is shown in Eq. (1):

$$x_{scaled} = \frac{(x-x_{min})}{(x-x_{max})} \tag{1}$$

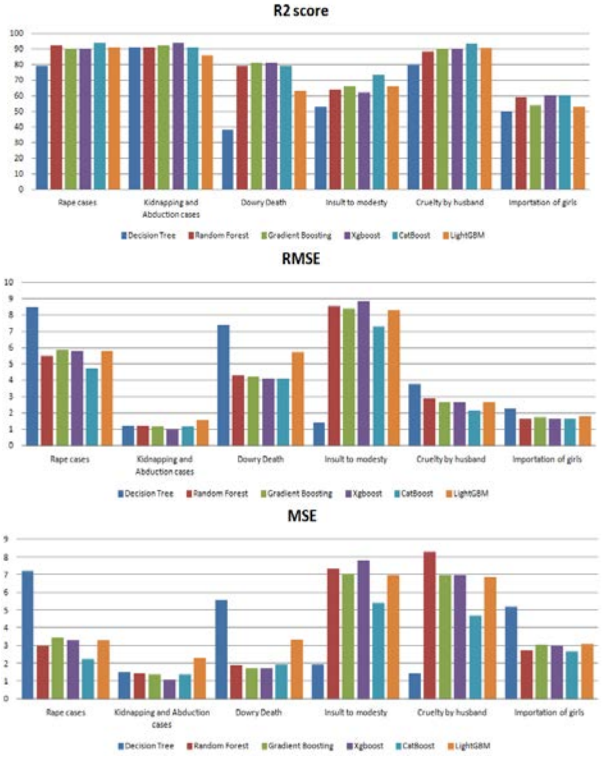


Fig. 10. Analysis of Algorithms using Min-Max Scalar.

On assaying Fig. 10, it can be said CatBoost worked well for importaton of girls, cruelty by husband, insult to modesty, and rape cases by 60, 93.4, 73.4, and 94 R^2 score, respectively while as XgBoost obtained great score in dowry death cases and kidnapping by 81 and 94 R^2 score respectively. On the other hand, for both mean square error and root mean square error best values have been obtained by CatBoost and XgBoost as compared to the other algorithms.

1) Standard scalar: It scales the values in such a way where the standard deviation or variance is 1 and mean is 0. The formula is shown in Eq. (2):

$$x_{scaled} = \frac{(x - mean)}{std\_dev} \qquad (2)$$

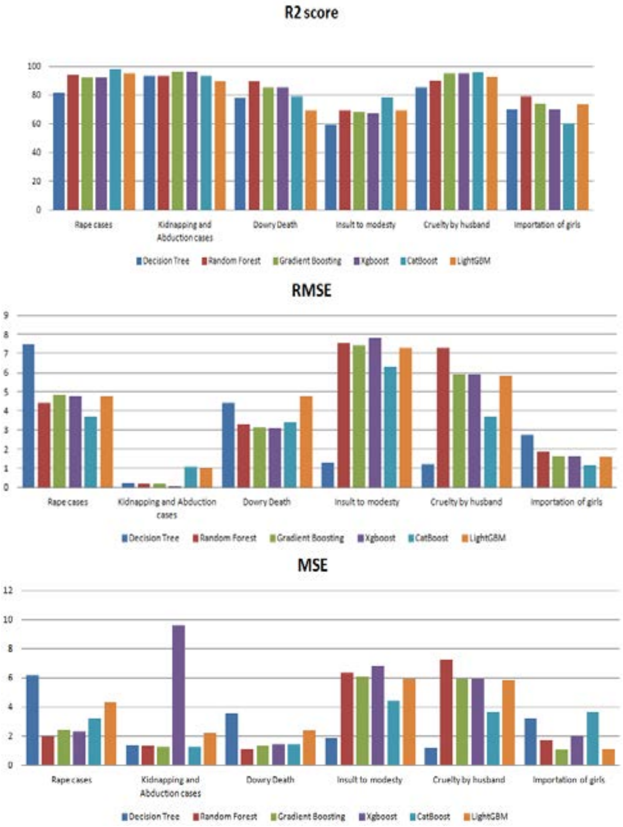


Fig. 11. Analysis of Algorithms using Standard Scalar.

The analysis in Fig. 11 determines that CatBoost worked well for the importation of girls, cruelty by husbands, an insult to modesty, and rape cases with 79, 93.4, 78.4, and 98 R2 scores, respectively. At the same time, XgBoost obtained a great score in dowry death cases and kidnapping with 89 and 96 R2 scores, respectively. On the other hand, for root mean square error and mean square error, the best values have been obtained by CatBoost and XgBoost as compared to the other algorithms.

2) Principal component analysis: A statistical technique in determining interrelations among a set of variables. The conversion of correlated variables to a set of uncorrelated variables takes place.

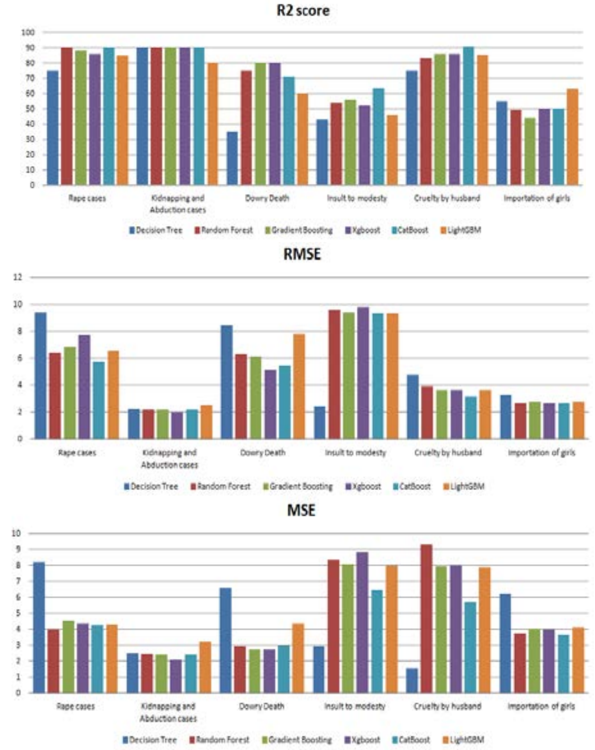


Fig. 12. Analysis of Algorithms using Principal Component Analysis.

On assaying Fig. 13, it can be said that LightGBM, CatBoost, XgBoost, and Random Forest gave better results for the women crime dataset by 63, 90.4, 80, 90 R2 scores, 2.76, 31.6,20.3,64.6 root mean square error value for data such as importation of girls, cruelty by husband, dowry, and rape, respectively.

The Scaling technique that gave the best results after normalizing the data was Standard Scalar. After feature scaling, the dataset has been divided into two halves. The first half consists of 75% of the training dataset, and the remaining half has a 25% testing dataset on which application of machine learning algorithms has been made.

F. *Model Selection*

Various machine learning models have been applied on the women crime dataset. A variety of models have been used on the dataset as one particular model cannot give accurate results as each model has its own characteristics and working technique. The output produced by each one of them is compared to get the best results. Features of each one of them have been discussed subsequently.

1) Decision tree: It is a representation of every all probable solutions to a problem/decision which is based on particular specifications [20]. It is organized in a tree-like form with two nodes: the Leaf Node and the Decision Node and Leaf nodes keep track of the outcomes of those decisions and have no other branches, on the other hand Decision nodes can have multiple branches depending upon the number of decisions made [21]. The features of the given dataset are used to make judgments or run tests. The values of decision tree can be calculated by computing the impurity of a node and its entropy which are shown in Eq. (3) and Eq. (4):

$$I_G(\mathrm{n}) = 1 - \sum_{i-1}^{J} (pi)^2 \qquad (3)$$

where J is the count of classes present in the node and p is the distribution of the class in the node.

$$Entropy = \sum_{i=1}^{c} -(pi)^* log_2(pi) \qquad (4)$$

where C is the number of classes present in the node and p is the distribution of the class in the node.

2) XgBoost: It is based on the concept of gradient boosting. It also makes use of decision trees. It is a custom, parallelized tree building algorithm which contains several decision trees. It provides features like efficient handling of missing data, and automatic feature selection [18].

3) Gradient boosting: Gradient Boosting is one of the most powerful techniques to construct predictive models. It trains many models in parallel. Every new model minimizes the loss function using the Gradient Descent Method. It is a greedy algorithm, and overfitting of the training dataset can happen quickly. It benefits from regularization methods by penalizing various parts of the algorithm and improving the algorithm's performance by reducing overfitting [22].

4) Random forest: It uses the concept of many decision trees to solve a compounded problem and helps improve the model's performance [23]. A random forest has the following characteristics. It combines several decision trees from distinct subsets of the provided dataset and averages them to increase the dataset's predicting accuracy. The number of trees and precision has a linear relationship, which helps to avoid over-fitting. It is divided into two phases: In the first phase, a random forest is created by mixing N decision trees, and in the second phase, predictions are made for every tree formed in the former step. The Random Forest algorithm is solved by Eq. (5) to Eq. (8):

$$RFfi_{\mathrm{i}} = \frac{\sum_{\mathrm{j}eall\,treas} normfi_{ij}}{T} \qquad (5)$$

$$normfi_{\mathrm{i}} = \frac{fi_i}{\sum_{jeall\,features} fi_j} \qquad (6)$$

$$fi_{\mathrm{i}} = \frac{\sum_{\mathrm{jinode\ j\ splits\ on\ feature\ i}} ni_j}{\sum_{keallnodes} ni_k} \qquad (7)$$

$$Ni_j = W_j C_j - W_{left(j)} C_{left(j) - W_{right(j)} C_{right(j)}} \qquad (8)$$

Here ni_{i} means importance of node j, $\mathrm{W_i}$ = weighted number of samples reaching node j, $\mathrm{C_i}$ = the impurity value of node j, left(j)= child node from left split on node j, right(j) = child node from right split on node j, fi_{i} = the importance of feature i, $RFfi_{\mathrm{i}}$ = the importance of feature i calculated from all trees in the Random Forest model, *normfi* = the normalized feature importance for i in tree j, T = total number of trees [19].

5) CatBoost: CatBoost is an open-source algorithm which uses gradient boosting on decision trees. It allows the use of non-numerical. It uses a combination of one-hot encoding and an advanced mean encoding [24].

6) LightGBM: Microsoft created LightGBM, a free and open-source distributed gradient boosting platform for machine learning. It's a decision tree-based gradient boosting framework that improves a model's efficiency while minimizing the utilization of memory. It uses a split strategy which is leaf-wise rather than level-wise, and consequently produces complex trees, which is the primary factor in attaining higher levels of accuracy [25].

G. Evaluation Metrics

Various evaluation metrics have been used such as accuracy; F1 score, precision, recall, root mean square error, and R score to test the performance of the applied machine learning models. Each one of them has been discussed briefly below:

Accuracy: The percentage of predicted values that match actual values is calculated by accuracy [25] and is shown in Eq. (9).

$$Acc = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ negative + False\ Positive + False\ Negative} \qquad (9)$$

F1 Score: It is used to assess the authenticity of a test. The Harmonic Mean of precision and recall is the F1 Score which has ranges between 0 and 1. It tells you how accurate and how robust our classifier is [26]. It is calculated by Eq. (10).

$$F1\ Score = 2\frac{Precision * Recall}{Recall + Precision} \qquad (10)$$

Precision: It is the number of positive classifiers divided by the total number of positives both true as well as false number of correct positive outcomes. [25]. It is calculated by Eq. (11).

$$Precision = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive} \qquad (11)$$

Recall: The number of rational positive results divided by total relevant samples [25]. It is calculated by Eq. (12).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \qquad (12)$$

Root mean square error: Euclidean distance uses to indicate how far predictions differ from true measured values [27]. It is one of the most widely used criteria for assessing the accuracy of predictions [28]. The root mean square error is calculated by Eq. (13).

$$RMSE = \sqrt{\sum_{i=1}^{j} \frac{II_{x(j)-x(j)} II^2}{n}} \qquad (13)$$

where n denotes the number of data points, x(j) refers to the j-th measurement, and x(j) belongs to the prediction which corresponds to the measurement.

R^2 score: It is known as coefficient of determination. It is the proportion of the variance in the dependent variable that is forecasted from the independent variable(s) [13] and is calculate by Eq. (14).

$$R^2 = 1 - \frac{SS_{RES}}{SS_{err}} \qquad (15)$$

where $\underline{SS_{RES}}$ is the residual sum of squared errors and SS_{err} is the total sum of squared errors.

IV. Results

The performance of the algorithms based on the various parameters such as accuracy, recall, root mean square error, precision, and F1 score has been computed during testing phase as shown in Table IV along with the discussion.

TABLE IV. Computation of Various Metrics for Machine Learning Models

Algorithms	**Metrics**				
	Accuracy	***Recall***	***RMSE***	***Precision***	***F1 Score***
Decision Tree	72	71	2.31	77	71
Random Forest	**92**	**85**	**0.91**	**86**	**85**
Gradient Boosting	74	80	2.25	79	80
Xgboost	88	78	1.35	78	78
CatBoost	79	84	2.11	84	85
LightGM	80	84	1.98	85	85

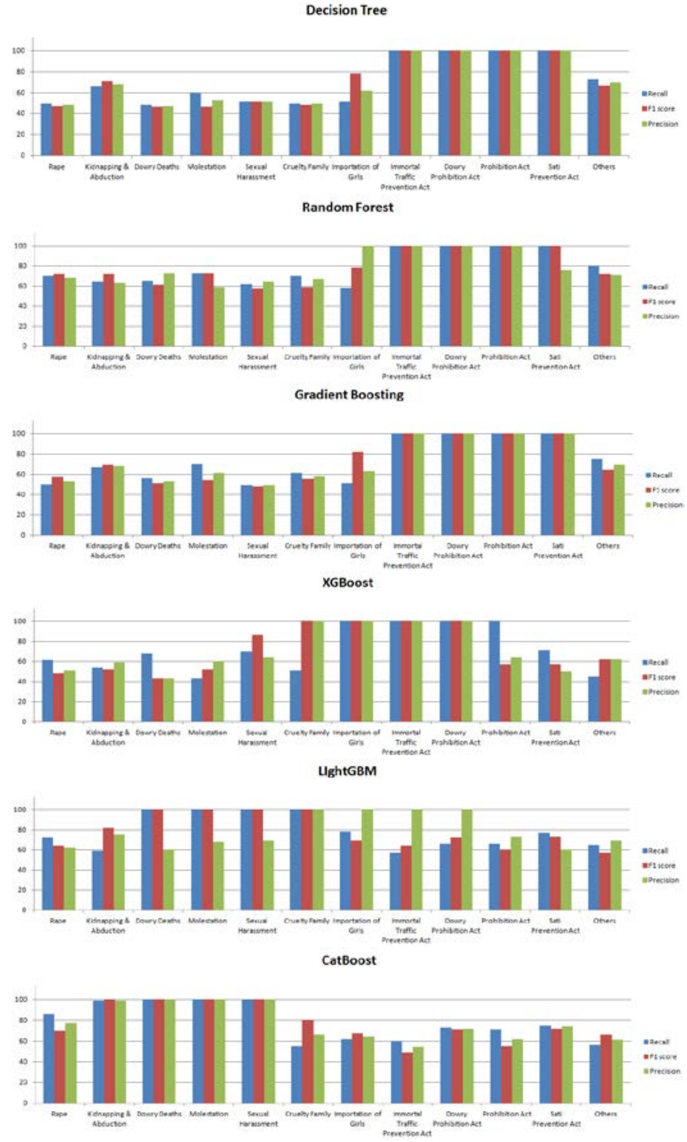


Fig. 13. Evaluation of Various Machine Learning Models using Three Different Metrics

On analyzing it can be said that random forest obtained the highest accuracy by 92%, recall by 85% and precision by 86%, F1 score by 85% on comparison with other algorithms. Besides this, each woman's recall, precision, and F1 score based on criminal activities such as dowry deaths, sexual harassment, importation of girls, prohibition act, sati prevention act, etc., have also been computed shown in Fig. 13. These parameters have been calculated to analyze the work of each learning model, as mentioned.

V. Discussion

In this research paper, the proposed system checks the credibility of the women based crime data by applying various machine learning models. Their performances have been evaluated using various evaluation metrics such as accuracy, recall, precision, F1 score, and root mean error square, with random forest achieving the highest accuracy of 92%, recall of 85%, the precision of 86%, F1 score of 85%, and the root mean square error value of 0.91 is obtained which is highest when compared to other algorithms. In addition to this, the work of various researchers in analyzing women's crime using different datasets has been also considered and compared with our study to understand our research work in a more efficient manner as shown in Table V.

TABLE V. Comparative Analysis of Previous Work with our Work

Ref	**Techniques**	**Dataset**	**Accuracy (%)**
[29]	Random forest	Crime in India dataset	86
[30]	Linear regression	National Crime Records Bureau (NCRB) crime data	83
[31]	Naïve Bayes	Collection of data from 2001 to 2012	81
[32]	KNN	Primary data	77
Our study	**Random forest**	**National Crime Record Bureau from year 2001 to 2019**	**92**

After comparing, it has been concluded that our study has achieved a great result in terms of accuracy for the National crime record bureau (2001-2019) with 92%, while linear regression has obtained less accuracy by 83% on applying the data that has been collected from the same repository. Overall, KNN has obtained the lowest accuracy value by 77% while working on the primary data.

VI. Challenges Faced during Research Work

During this research, we faced specific challenges while collecting the data, which is the most important step for crime-based research. Like this, the three main problems that we came across were:

1) No uniformity in Data: The data collected initially from 2001 to 2015 had a different data format and many parameters related to the case status. Still, the data from 2015 onwards was more compressed and less comprehensive. In addition, many crimes go unreported, and no one has a record of them.

2) Varying data formats: The data from 2001 to 2016 was in excel format and had a different format and parameters, and data from 2016 onwards was in pdf format and had a different set of parameters. A lot of time was taken to pre-process data into a desired, standardized format.

3) Size of the crime data: The data was huge as there are 11 sub-categories of crime against women, and each crime sub-category case status is stored. There are 13 types of case status. All this information is stored for 27 states and 8 Union Territories. The records taken were for 15 years. As the data size was huge, rectifying the problems with the data was an uphill task. Along with it, the data taken from books was compatible for analysis.

4) Missing values in crime data: There were a lot of missing values in the data, so the data cleaning was time-consuming. Hence, the SimpleImputer technique was used to handle missing data.

VII. Conclusion

The study focuses on the investigation of women's crime statistics such as rape, abduction, dowry deaths, molestation, sexual harassment, cruelty to family, importation of girls, immortal traffic, dowry prohibition act, prohibition act, sati prevention act, and others from 2001 to 2019 in various Indian states using data provenance and machine learning models. It has been discovered that Random Forest proved to be the best approach when compared to the others. Moreover, the proposed system demonstrates that the algorithms used for machine learning perform effectively in analyzing various crimes against women across states and union territories. Hence, the same technique can also be applied to search for information about other crimes in states with a higher crime rate. In addition, deep learning algorithms might be utilized to improve prediction accuracy while dealing with complex criminal cases in India.

References

[1] Das, Priyanka, and Asit Kumar Das. "Crime analysis against women from online newspaper reports and an approach to apply it in dynamic environment." 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC). IEEE, 2017.

[2] Hagan, Frank E. Crime types and criminals. Sage, 2009.

[3] Balasubramanian, t. “Violence against women’s in India”. Journal of shanghai jiaotong university, 876-892, 2020.

[4] Saravanan, Parthasarathy, et al. "Survey on crime analysis and prediction using data mining and machine learning techniques." Advances in Smart Grid Technology. Springer, Singapore, 2021. 435-448.

[5] Mittal, Mamta, et al. "Monitoring the impact of economic crisis on crime in India using machine learning." Computational Economics 53.4 (2019): 1467-1485.

[6] Shah, Neil, Nandish Bhagat, and Manan Shah. "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention." Visual Computing for Industry, Biomedicine, and Art 4.1 (2021): 1-14.

[7] Braga, Anthony A., et al. "Hot spots policing of small geographic areas effects on crime." Campbell Systematic Reviews 15.3 (2019): e1046.

[8] Nakib, Mohammad, et al. "Crime scene prediction by detecting threatening objects using convolutional neural network." 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2). IEEE, 2018.

[9] Englbrecht, Ludwig, et al. "Enhancing credibility of digital evidence through provenance-based incident response handling." Proceedings of the 14th International Conference on Availability, Reliability and Security. 2019.

[10] Telugu M.,Vaddemani Sai M., K V Sai S. ,G. Shriphad R. Crime Data Analysis Using Machine Learning Models”, IJAST, vol. 29, no. 9s, pp. 3260 – 3268, 2020.

[11] Tamilarasi, P., and R. Uma Rani. "Diagnosis of crime rate against women using k-fold cross validation through machine learning." 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2020.

[12] Anjali, “ Women empowerment and constitutional provisions”. In Legalserviceindia.com.

[13] Qi, Zhang. "The text classification of theft crime based on TF-IDF and XGBoost model." 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE, 2020.

[14] Shermila, A. Mary, Amrith Basil Bellarmine, and Nirmala Santiago. "Crime data analysis and prediction of perpetrator identity using machine learning approach." 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2018.

[15] Amponsah, Wellington, and Parvinder Kaur. "Exploratory Data Analysis And Crime Prevention Using Machine Learning: The case of Ghana.".

[16] Mishra, Shivani, and Suraj Kumar. A comparative study of crimes against women based on Machine Learning using Big Data techniques. No. 4376. EasyChair, 2020.

[17] Tamilarasi, P., and R. Uma Rani. "Predict the Crime Rate Against Women Using Machine Learning Classification Techniques." Data Science and Its Applications. Chapman and Hall/CRC, 2021. 295-313.

[18] Durgapal, Vartika Hari. "Crime Based Evaluation of GPS Network Using Machine Learning Techniques." 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2022.

[19] Prabakaran, S., and Shilpa Mitra. "Survey of analysis of crime detection techniques using data mining and machine learning." Journal of Physics: Conference Series. Vol. 1000. No. 1. IOP Publishing, 2018.

[20] Ivan, Niyonzima, et al. "Crime Prediction Using Decision Tree (J48) Classification Algorithm." (2017).

[21] Patel, Nisarg P., et al. "Fusion in Cryptocurrency Price Prediction: A Decade Survey on Recent Advancements, Architecture, and Potential Future Directions." IEEE Access 10 (2022): 34511-34538.

[22] Ghankutkar, Surabhi, et al. "Modelling machine learning for analysing crime news." 2019 International Conference on Advances in Computing, Communication and Control (ICAC3). IEEE, 2019.

[23] Mistry, Chinmay, et al. "MedBlock: An AI-enabled and blockchain-driven medical healthcare system for COVID-19." ICC 2021-IEEE International Conference on Communications. IEEE, 2021.

[24] Kumar, Yogesh, Komalpreet Kaur, and Gurpreet Singh. "Machine learning aspects and its applications towards different research areas." 2020 International conference on computation, automation and knowledge management (ICCAKM). IEEE, 2020.

[25] Rawat, Romil, et al. "Analysis of darknet traffic for criminal activities detection using TF-IDF and light gradient boosted machine learning algorithm." Innovations in electrical and electronic engineering. Springer, Singapore, 2021. 671-681.

[26] Sachin Bhardwaj Yogesh Kumar, M. K. P. K. R. Recent Trends of Data Mining in the Field of Intrusion Detection System. Journal of Critical Reviews, 7, 2360–2365,2020.

[27] Parekh, Raj, et al. "DL-GuesS: Deep Learning and Sentiment Analysis-based Cryptocurrency Price Prediction." IEEE Access 10 (2022): 35398-35409.

[28] McClendon, Lawrence, and Natarajan Meghanathan. "Using machine learning algorithms to analyze crime data." Machine Learning and Applications: An International Journal (MLAIJ) 2.1 (2015): 1-12.

[29] Sarah, J., Danny, A., Deen, J., Dongre, L., Chitransh, V. and Ramchandhani, H. Analysing Crimes of Indian Datasets Based on Machine Learning Methods. In vol 12, issue 8, 2415-2435, 2021.

[30] Ravi Teja, K., et al. "Analysis of Crimes against Women in India Using Machine Learning Techniques." Communication Software and Networks. Springer, Singapore, 2021. 499-510.

[31] Prasad, D. S., Rachit Sharma, and V. Anbarasu. "Analysis and Prediction of Crime against Woman Using Machine Learning Techniques." Annals of the Romanian Society for Cell Biology 25.6 (2021): 5183-5188.

[32] Mahmud, Sakib, Musfika Nuha, and Abdus Sattar. "Crime rate prediction using machine learning and data mining." Soft Computing Techniques and Applications. Springer, Singapore, 2021. 59-69.

[33] Sonal Singh. “Leveraging ML to Predict Crime Against Women” International Journal of Engineering Research & Technology (IJERT) ,Volume 11, Issue 01 (January 2022),

[34] Stoffel, Florian, et al. "VAPD: A Visionary System for Uncertainty Aware Decision Making in Crime Analysis." Symposium on Visualization for Decision Making Under Uncertainty at IEEE VIS 2015. 2015.

[35] https://ncrb.gov.in/crimeinindiatablecontents?field_date_value[value][year]=2019&field_select_additional_table_ti_value=All&items_per_page=All.

[36] https://www.kaggle.com/datasets/khanmohammadanas/district-wise-crimes-in-india.

XBLQPS: An Extended Bengali Language Query Processing System for e-Healthcare Domain

Kailash Pati Mandal[1], Prasenjit Mukherjee[2], Atanu Chattopadhyay[3], Baisakhi Chakraborty[4]
Computer Science and Engineering, National Institute of Technology, Durgapur, India[1, 2, 4]
BBA (H) and BCA (H) Department, Deshabandhu Mahavidyalaya, Chittaranjan, India[3]

***Abstract*—The digital India program encourages Indian citizens to become conversant with e-services which are primarily English language-based services. However, the vast majority of the Indian population is comfortable with vernacular languages like Bengali, Assamese, Hindi, etc. The rural villagers are not able to interact with the Relational Database Management system in their native language. Therefore, create a system that produces SQL queries from natural language queries in Bengali, containing ambiguous words. This paper proposes a Bengali Query Processor named Extended Bengali language Query Processing System (XBLQPS) to handle queries containing ambiguous words posted to a Healthcare Information database in the electronic domain. The Healthcare Information database contains doctor, hospital and department details in the Bengali language. The proposed system provides support for the Bengali-speaking Indian rural population to efficiently fetch required information from the database. The proposed system extracts the Bengali root word by removing the inflectional part and categorizing them to a specific part of speech (POS) using modified Bengali WordNet. The proposed system uses manually annotated parts of speech detection of a word based on Bengali WordNet. Patterns of noun phrases are generated to detect the correct noun phrase as well as entity and attribute(s). Entity and attributes are used to prepare the semantic table which is utilized to create the Structured Query Language (SQL). The simplified LESK method is utilized to resolve ambiguous Bengali phrases in this query processing system. The accuracy, precision, recall and F1 score of the system is measured as 70%, 74%, 73%, and 73% respectively.**

Keywords*—Relational database management system (RDBMS); modified bengali WordNet; LESK algorithm; structured query language (SQL); natural language query

I. Introduction

Indian citizens have become more familiar and conversant with online or electronic systems like e-banking, e-governance, e-health, e-tourism, and e-education due to the empowerment through the digital India program. Nowadays most government facilities are electronic based. However, the vast majority of the Indian population is village based. There are 0.7 million villages in India. Most villagers are not accustomed to the English-based systems. They are comfortable with their own vernacular language(s) and Bengali is one of them. Bengali is widely used in West Bengal, Andaman and Nicobar Islands, Tripura, Assam, and other states. The Bengali language is the official language of Bangladesh, West Bengal and Tripura. One of the major e-service requirements in the rural interiors is related to healthcare systems. Health care domain-based e-services in vernacular languages are a really challenging task. Today, the query-response model in vernacular language is an open research problem in the research community. This is because the resources of vernacular languages are very low and difficult to implement. A good query response model in the Bengali language can be helpful for naive users to extract information without any technical knowledge. WordNet plays a vital role to develop the query-response model. The Query expansion (QE) is a well-known technique used to enhance the effectiveness of information retrieval. A new approach for QE using Wikipedia and WordNet as data sources is described in [1]. The IndoWordNet is an important lexicographic resource for different Indian languages. The lexical matrix has been used to construct the relational semantics as discussed in [2]. In the article [3], The BWN is a WordNet database for the Bengali language. It consists of lexical source files, a grinder, a WordNet database, and an interface. Using the WNDB interface, users can interact with the BWN documents in various ways as explained in [3]. The African Wordnet Project aims to develop aligned wordnets for African languages spoken in South Africa. The focus of this article will be on isiZulu as one of the selected languages used in [4]. The meaning of an ambiguous word has been determined according to its context. The proposed method generates context by comparing a doubtful word with words in the input document for similarity. The similarity computation is based on BabelNet's semantic framework. [5].

The naive users who access the online Healthcare Information Management System in the Bengali language is a challenging task. Therefore, this paper proposes a modified Bengali WordNet and a natural language query-response system, namely Extended Bengali language Query Processing System (XBLQPS) that can handle natural language queries in Bengali. This automated system generates responses from the knowledge database. A relational Database Management System (RDBMS) is used to implement the knowledge database. This work is an enhancement of the work of [6], where a Query Processing System for the Bengali language has been proposed and discussed.

This research article consists of various sections. Section II narrates related works. Section III describes the objective of the proposed system where Section IV explains the proposed system and Section V describes analysis and discussion of the proposed system. Section VI is about the modified Bengali WordNet and health information database. Section VII illustrates the used tool and IPA notation. Section VIII derives the time complexity of the proposed system. Section IX is related to analyze the precision, recall, and accuracy of the

XBLQPS, Section X presents the limitations and future works. The proposed system is concluded in Section XI.

II. Related Work

Bengali WordNet databases are useful and a few good implementations have been done by Pushpak Bhattacharyya [2] and Farhana Farque [3]. The proposed work is influenced from Pushpak Bhattacharyya [2], Farhana Farque [3] and A. Haque [22].

The article [7] stated a question-answering model using natural language query where interactions between table and question are complex type. The sketch-based approach has been applied to solve the complexity as in [7]. Word sense disambiguation (WSD) is to find the proper meaning of a word in any context. It involves incorporating word knowledge from external knowledge resources to remove equivocalness. A WSD tool has been implemented using Hindi WordNet where it is a knowledge-based approach as in [8]. In the paper [9], the sense Induction approach has been used in the algorithm for word sense disambiguation (WSD) in Bengali. Ten frequently used Bengali ambiguous terms and 200 phrases of each term are utilized to test the WSD model. The radix tree-based structure is used to keep context information as well as faster searching of a word in the paragraph. The WordNet is used as a knowledge base to disambiguate the word as discussed in [10]. The Case-Based Reasoning interpretation technique was used to decipher the confusing word Gurmukhi, also known as Punjabi, in Indian Regional Language. The solution to the new problem was inferred from the previously solved problems as described in [11]. Machine translation at the human level can be helpful greatly using WSD. Through the FP-Growth method, Authors [12] provide a system for WSD in their study as in [12]. The fuzziness of semantic relation has been applied in Fuzzy Hindi WordNet (FHWN). Various membership values of semantic relations of the FHWN were considered to extract the correct sense as in [13]. In the article [14], the authors described the root word extraction technique from the Bengali inflected nouns by applying well-defined grammatical mapping rules (GMRs) between nominal bases and inflections. The proposed methodology can be applied to the Bengali grammar and inflected words that were described in [14]. To decipher the confusing term, the LESK algorithm has been utilized. The result set of the proposed system is in line with the result set of KBBI as well as provides an accuracy of 78.6% for one of the ambiguous words while 62.5 % for two ambiguous words as explained in [15]. The effort of identifying the meaning of a word in a certain situation is known as word sense disambiguation (WSD). The novel WSD [16] model has been introduced where the proposed model calculates each word's meaning uniquely. The creation of Arabic WordNet (AWN) has made lexical resources available to the Arabic NLP community. The usage of this resource cannot be considered because there are fewer AWN Synsets than other WordNets that has been elaborated in [17]. In the article [18], Vietnamese and Korean both are morphologically rich languages. The high homograph rate in the Korean language is word ambiguities that affect neural MT (NMT). There isn't a sufficient, publicly accessible parallel Korean and Vietnamese that can be utilized to train translation models as implemented in [18]. A hybrid approach is used for Urdu word stemming that is helpful for information extraction, textual categorization, data analysis and related applications. This proposed approach [19] works on unigram, bigram, and trigram features that were discussed in [19]. The word was disambiguated by combining the sense relativeness algorithms with a neural model in the paper [20]. The proposed model works on POS-labeled text corpus and the length of the context may be varied as discussed in [20]. The raw collection of Bangla text has been used to generate meaning-tagged data. The Bangla meaning tagged data contains root word form and their POS type of an ambiguous word with 86.95% performance as implemented in [21]. The Bangla Word Sense Disambiguation System can be distinguished by some confusing Bangla phrases. Parsing and detection are two main working phases of the proposed system [22]. An Algorithm has been developed that clears the confusing word according to the categories of ambiguous words that are nouns, adjectives, and verbs as in [22]. Authors were focused on how to differentiate the important records and generate a summary from them. In the proposed system, the authors applied natural language processing, WordNet and lexical chains for the summary generation of a text as explained in [25]. An Arabic WordNet (AWN) has been used to overcome the WSD problem where word semantic similarity has been checked by multiple Arabic stemming algorithms. This work is related to reducing the gap in Arabic NLP compared to English as in [26]. Table I shows the results of a comparative investigation of existing systems with the proposed system.

TABLE I. A Comparison of Identical Systems using the XBLQPS

SL NO.	AUTHOR(S) & SYSTEM	METHODOLOGY USED IN SAME TYPE SYSTEM	METHODOLOGY APPLIED IN THE PROPOSED SYSTEM (XBLQPS)
1	S. Basuki et al. & LESK Algorithm Utilization for Word Sense Disambiguation (WSD) for Indonesian Homograph Word Meaning Determination [15]	1) The LESK procedure has been used to disambiguate Indonesian Homograph Word referred in [15]. 2) This system provides 78.6% accuracy if one ambiguous word present, and 62.5% accuracy if two ambiguous words present. 3) The time complexity has not been discussed for this system	1) The proposed system uses the LESK algorithm, modified Bengali WordNet and formation of patterns to disambiguate the Bengali word. 2) The overall accuracy of XBLQPS is 70%. 3) The time complexity has been calculated for the proposed system.
2	M. S. Kaysar et al.& Applying FP-Growth Algorithm to Disambiguate Bengali Words [12]	1) The FP-Growth Algorithm has been used to disambiguate the Bengali ambiguous word. 2) This system provides 80% accuracy. 3) The time complexity has not been specified here.	1) The LESK has been used to disambiguate the Bengali ambiguous word. 2) The XBLQPS provides 70% accuracy. 3) The complexity has been computed here.

3	D. O et al. & Word Sense Disambiguation Utilizing Word Vector Representation from a Knowledge-based Graph Based on Word Similarity Calculation [5]	1) The system referred in [5] has been used to disambiguate the ambiguous word using similarity calculation with help of semantic network structure of BabelNet. 2) This system does not generate pattern(s). 3) The precision, recall and F1 score on semEval-2013 dataset are 75%, 75%, and 75% respectively whereas on semEval-2015 are 69.2%, 62.6% and 65.8%.	1) The XBLQPS uses LESK algorithm and modified Bengali WordNet to disambiguate the ambiguous word. 2) The XBLQPS generates pattern(s). 3) The precision, recall and F1 score for the proposed system are 74%, 73% and 73% respectively.
4	M. Biswas et al. & Construction of a Bangla Sense Annotated Corpus for Disambiguation of Word Sense [21]	1) The system referred in [21] creates sense annotated corpus containing ambiguous word. 2) The accuracy of sense annotated corpus is 86.95%. 3) The time complexity has not been discussed for this system.	1) The XBLQPS creates SQL from the Bengali language query containing ambiguous word. 2) The proposed system provides 70% accuracy. 3) The time complexity of XBLQPS has been computed.
5	K. P. Mandal et al. & An unique Bengali Language Query Execution System in the field of health [6]	1) The system referred in [6] does not able to handle the query containing ambiguous word. 2) This system does not use LESK algorithm. 3) The time complexity of this system is $O(n^4)$.	1) The XBLQPS can process the query containing ambiguous word. 2) This system uses LESK algorithm. 3) The time complexity of the proposed system is $O(n^3)$.
6	C. Lachichi et al. & Machine translation and external linguistic elements have been used to enrich Arabic WordNet[17]	1) The system referred in [17] uses Machine Translation and External Linguistic Resources to enhance the Arabic WordNet. 2) It is an Arabic WordNet enhancement system. 3) The average accuracy of this system is 0.48.	1) The XBLQPS uses Bengali WordNet to disambiguate the Bengali ambiguous word. 2) It is a Bengali language query processing system. 3) The accuracy of the XBLQPS is 0.70.
7	A. Haque et al. & Utilizing dictionary-based approach, a Bangla word sense disambiguation model has been proposed [22]	1) The system referred in [22] is a dictionary based Bengali word sense disambiguation system. 2) The accuracy of this system is 82.40%. 3) This system does not convert Bengali natural query to SQL.	1) The XBLQPS is a query processing system containing Bengali ambiguous word. 2) The accuracy of the XBLQPS is 70%. 3) The XBLQPS converts Bengali natural language query to SQL.

III. Objective of the Proposed System

- The main objective of this research is to create a system that can handle natural language queries in Bengali in the healthcare domain. A naïve user will be able to extract healthcare information without any technical knowledge.
- The proposed system will be able to handle natural language queries that are containing ambiguous words in Bengali.
- The proposed system will be able to generate a response from the Bengali Query that contains ambiguous words.
- The proposed system will generate SQL from natural language queries in Bengali without any manual intervention.

IV. Proposed XBLQPS

Following are the steps of the proposed system:

Step 1: The user logs into the proposed system and submits the Bengali language query.

Step 2: The system slices the query into token(s).

Step 3: The root word extraction and POS tagging are done with the help of modified Bengali WordNet.

Step 4: Once the POS tagging is over, the proposed system generates pattern(s) with the help of noun and noun phrases.

Step 5: The ambiguous term is disambiguated using the simplified LESK method. The proper sense of an ambiguous Bengali word was discovered by examining the most significant number(s) of common words existent between the word's current context and gloss.

Step 6: After identification of the correct sense of every pattern, entity and attribute are also determined from the table named “table_entity_attribute_sensing” for semantic analysis.

Step 7: Once entity and attribute are detected, a set of predefined rules are used to generate an SQL query.

Step 8: After generating the SQL query, the proposed system executes and retrieves desired information from the health information database.

There are three tables in the health information database. "table_hospital" contains the hospital's name and address. The "table_doctor" includes the doctor's name, qualification, specialization, fee, and a particular doctor connected with the department and hospital. The “table_department” contains the department’s name and which department exists in which hospital.

Finally, the desired result sends to the user. The workflow diagram of the XBLQPS is depicted in Fig. 1. The workflow diagram (Fig. 1) shows the functionality of each component of the proposed system.

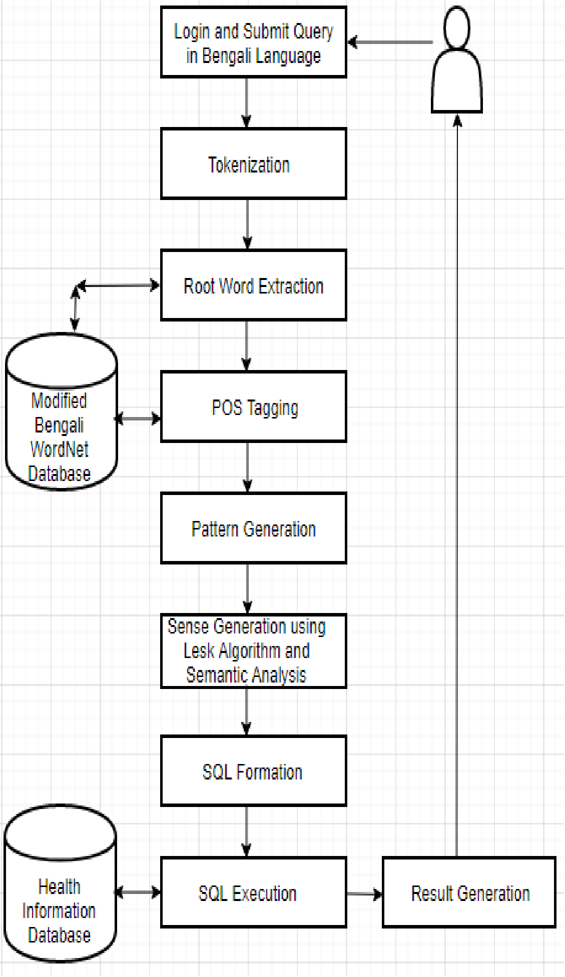

Login and Submit Query in Bengali Language
Tokenization
Root Word Extraction
Modified Bengali WordNet Database
POS Tagging
Pattern Generation
Sense Generation using Lesk Algorithm and Semantic Analysis
SQL Formation
Health Information Database
SQL Execution
Result Generation

Fig. 1. Work Flow Diagram of XBLQPS.

V. Analysis and Discussion of the Proposed System

The detailed analysis and description of each component of the proposed system are as follows.

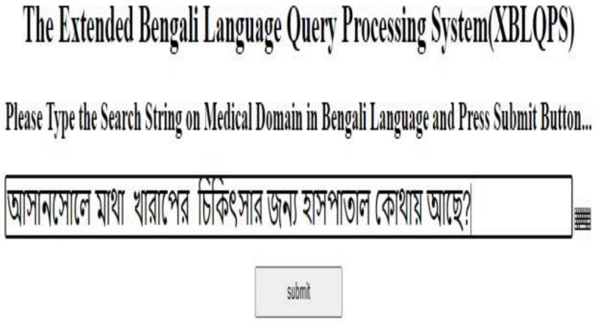
Component 1: Login into the system and submit the query in the Bengali language.

The user logs in to the proposed system and submits the query string in the Bengali language. For instance, a query has been provided. The example of query 1 is given in Fig. 2.

আসানসোলে মাথা খারাপের চিকিৎসার জন্য হাসপাতাল কোথায় আছে?

(asansole matʰa kʰaraper tʃikitsar dʒno ɦaspatal kotʰae̯ atʃʰe?)

(Where is a hospital for treatment of mental diseases in Asansol?)



The Extended Bengali Language Query Processing System(XBLQPS)

Please Type the Search String on Medical Domain in Bengali Language and Press Submit Button...

আসানসোলে মাথা খারাপের চিকিৎসার জন্য হাসপাতাল কোথায় আছে?

Fig. 2. Example Query 1 in Bengali.

Component 2: Tokenization

The query string will be sliced into small linguistic units called tokens after removing punctuation mark(s). These token(s) are stored in a string array. Table II shows the array indexing in the string array of the user-submitted query after tokenization.

Component 3: Root word extraction

The nominal word is the main backbone of SQL formation. Often the noun exists in the inflected form in the Bengali sentences as Bengali is a highly inflected language. Before POS tagging and extraction of correct sense, each token is compared with the inflectional part table i.e. “table_inflectional_part” which is shown in Table XV in the modified Bengali WordNet DB. The proposed system compares each token with the inflectional part table, that is, Table XV. If any of the inflectional parts in the inflectional part table matches with the trailing part of the token, then the XBLQPS removes the matched part from the token. In this way, the proposed system extracts the root word by removing the matched part from the token. If the trailing part of the token does not match with the inflectional part table in the Modified Bengali WordNet DB, then the XBLQPS will consider that the token itself will be a root word. After removing the inflectional part, all token(s) will be stored again in another string array. After extraction of the root word, the string array has been given in Table III.

TABLE II. Tokenization

array index	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
String array of user query	আসানসোলে (asansole)	মাথা (matʰa)	খারাপের (kʰaraper)	চিকিৎসার (tʃikitsar)	জন্য (dʒno)	হাসপাতাল (ɦaspatal)	কোথায় (kotʰae̯)	আছে (atʃʰe)

TABLE III. Root Word Extraction

array index	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
String array of root word	আসানসোল (asansol)	মাথা (matʰa)	খারাপ (kʰarap)	চিকিৎসা (tʃikitsa)	জন্য (dʒno)	হাসপাতাল (ɦaspatal)	কোথায় (kotʰae̯)	আছে (atʃʰe)

Component 4: POS tagging

The proposed system utilizes a POS string array (Table IV) with the same length as the string array of the root words (Table III) for keeping track of parts of speech of every token. The POS string array (Table IV) is helpful for pattern generation of noun phrases. The proposed system selects each token and compares it with the table named “table_word” (Table IX) in the modified Bengali WordNet DB. If the token matches with the table named “table_word” in the modified Bengali WordNet DB, then the corresponding “word_id” field value will be retrieved. The retrieved “word_id” will be compared with the “word_id” field of the table named “table_sense” is shown in Table XI, from where the proposed system will map the “synset_id”. Using “synset_id”, the XBLQPS will retrieve the “parts_of_speech” field value from the table “table_synset”is given in Table X and that will be inserted into the POS string array (Table IV). If the token does not match with the modified Bengali WordNet DB, then the token will be treated as noun (বিশেষ্য) and that will be inserted into the POS string array (Table IV) as a noun (বিশেষ্য). The selected token's array index value is same in POS string array (Table IV) and string array of root word (Table III). Each token of the query will be compared in the modified Bengali WordNet DB and on finding a match with a word, it will be stored in POS string array (Table IV) in the form of its parts of speech derived from the parts of speech definition given alongside the word in the modified Bengali WordNet proposed in the paper, except the token আসানসোল (Asansol) which is a proper noun; the name of a place. That why the POS string array (Table IV) value of আসানসোল is treated as a noun (বিশেষ্য), which means all unknown tokens which do not match with the modified Bengali WordNet DB will be treated as a noun (বিশেষ্য). The string array after pos tagging has been given in Table IV.

Component 5: Pattern generation of noun phrases

The noun and noun phrases are very much important for entity and attribute(s) identification which are main components of SQL formation. The correct noun or noun phrases have been identified by pattern generation because often the Bengali noun phrase is consisting of more than one consecutive noun (বিশেষ্য) or noun with an adjective (বিশেষণ). The proposed system will generate pattern(s) for the consecutive noun (বিশেষ্য) or consecutive any combination of a noun (বিশেষ্য) and adjective (বিশেষণ). More than one consecutive noun (বিশেষ্য) or combination of noun (বিশেষ্য) and adjective (বিশেষণ) will generate more than one pattern whereas nonconsecutive noun (বিশেষ্য) will generate single pattern. The token which is not belong the noun (বিশেষ্য) or adjective (বিশেষণ) is simply ignores by the proposed system. The system will generate more than one pattern for array index 0, 1, 2 and 3 of the POS string array (Table IV) because this array index contains combination of noun (বিশেষ্য) and adjective (বিশেষণ). The single pattern will generate for array index 5 because this is a non-consecutive noun (বিশেষ্য). During pattern generation, the token order occurrence will be maintained by the proposed system and that will be stored in a string array. The string array of pattern has been given in Table V. The generated patterns for the user given query are as follows.

1) আসানসোল (asansol)
2) মাথা (mat^{h}a)
3) খারাপ (k^{h}arap)
4) চিকিৎসা (tʃikitsa)
5) আসানসোল মাথা (asansol mat^{h}a)
6) মাথা খারাপ (mat^{h}a k^{h}arap)
7) খারাপ চিকিৎসা (k^{h}arap tʃikitsa)
8) আসানসোল মাথা খারাপ (asansol mat^{h}a k^{h}arap)
9) মাথা খারাপ চিকিৎসা (mat^{h}a k^{h}arap tʃikitsa)
*10)*আসানসোল মাথা খারাপ চিকিৎসা (asansol mat^{h}a k^{h}arap tʃikitsa)
*11)*হাসপাতাল (ɦaspatal)

Component 6: Sense extraction of pattern(s) using LESK algorithm and semantic analysis

After generation of pattern(s), every pattern will be compared with the value of “word_name” field of the table named “table_word” (Table IX) in the modified Bengali WordNet DB. If the pattern matches with the value of “word_name” field of the table named “table_word”, then corresponding “word_id” will be fetched. The fetched “word_id” will be compared with the value of “word_id” field of the table named “table_sense” is shown in Table XI. If the match is found then corresponding synset_id will be retrieved. Finally, the retrieved “synset_id” will be compared with the value of “synset_id” of the table named “table_synset” is indicated in Table X. If the match is found then the corresponding “parts_of_speech”, “gloss_concept_defination”, “example_sentence”, “meaning” and “possible_attribute” field’s values will be retrieved from the table named “table_synset”. The retrieved “possible_attribute” value will be compared with the value of “attribute_name” field of the table named “table_entity_attribute_sensing” is specified in Table XIV. If the“possible_attribute” field’s value of the table named “table_synset” matches with any value of “attribute_name” field of table named “table_entity_attribute_sensing”, then corresponding row value will be fetched from “table_entity_attribute_sensing”.

TABLE IV. POS TAGGING

Array Index	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
POS string array	বিশেষ্য (biʃeʃɔ)	বিশেষ্য (biʃeʃɔ)	বিশেষণ (biʃeʃn)	বিশেষ্য (biʃeʃɔ)	অব্যয় (oboe̯)	বিশেষ্য (biʃeʃɔ)	সর্বনাম (sr[2]bonam)	ক্রিয়া (kir[2]ia)

TABLE V. PATTERN GENERATION OF NOUN PHRASES

Array Index	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
String array of patter	আসানসোল (asansol)	মাথা (matʰa)	খারাপ (kʰarap)	চিকিৎসা (tʃikitsa)	আসানসোল মাথা (asansol matʰa)	মাথা খারাপ (matʰa kʰarap)	খারাপচিকিৎসা (kʰarap tʃikitsa)	আসানসোলমাথা খারাপ (asansol matʰakʰarap)	মাথা খারাপচিকিৎসা (matʰa kʰarap tʃikitsa)	আসানসোলমাথা খারাপ চিকিৎসা (asansol matʰa kʰarap tʃikitsa)	হাসপাতাল (ɦaspatal)

The row which will be fetched from "table_entity_attribute_sensing" that will be inserted into "table_semantic"is given in Table XVI for semantic analysis where the value of "meaning" field of the tablenamed "table_synset" will be inserted into the "value" field of the table named "table_semantic", If anytime exists NULL value in the "meaning" field then the "value" field of the table named "table_semantic" will also be NULL. Sometimes the pattern may contain more than one value at the "possible_attribue" field. The proposed system fetches all values of "possible_attribute" field and compared with value of "attribute_name" field of the table named "table_entity_attribute_sensing". The matched row values are fetched from "table_entity_attribute_sensing", inserted into "table_semantic" for semantic analysis. If the pattern contains NULL value at their "possible_attribue" field then the pattern simply ignores by the proposed system. If a "word_id" contains only one "synset_id" in the table named "table_sense", that means the pattern is unambiguous .If a"word_id" contains more than one "synset_id" in the table named "table_sense", that means the pattern is ambiguous. To identify the correct sense of the ambiguous pattern, the XBLQPS considers user given query as well as value of the "gloss_concept_defination" and "example_sentence" fields of the table named "table_synset". Next step, the system applies LESK algorithm referred in [27] to find out correct sense of ambiguous pattern(s). The XBLQPS counts the number of common word(s) by comparing between current user given query with the value of "gloss_concept_defination" field of the table named "table_synset" as well as the value of "example_sentence field". The proposed system adds up those above-mentioned count values and selects the row value from the table named "table_synset" which gives maximum count value. The pseudo code of the simplified LESK algorithm is shown in Table VI.

TABLE VI. THE PSEUDO CODE OF THE SIMPLIFIED LESK ALGORITHM

```
Function LESK_simple(word, sentence)
Ideal sense=widely used sense of a word
Maximum common word=0
Context=number of word present in the sentence
Do select every meaning from a set of meaning of every word
Favourite=a set of words present in the gloss example
Collision=compute_collision(favourite, context)
If collision > maximum common word then
Maximum common word=collision
Ideal=meaning
End [27].
```

Eleven patterns have been generated from the user given query; among them five patterns i.eমাথা (matʰa), খারাপ (kʰarap), চিকিৎসা (tʃikitsa), মাথা খারাপ (matʰa kʰarap) and হাসপাতাল (ɦaspatal) match with table named "table_word". The remaining six patterns i.e. আসানসোল (asansol), আসানসোল মাথা (asansol matʰa), খারাপ চিকিৎসা(kʰarap tʃikitsa), আসানসোল মাথা খারাপ (asansol matʰa kʰarap), মাথা খারাপ চিকিৎসা (matʰa kʰarap tʃikitsa) and আসানসোল মাথা খারাপ চিকিৎসা (asansol matʰa kʰarap tʃikitsa) do not match with table named "table_word" and the proposed system treats these mismatched pattern(s) as a value. These values will be compared with the three tables named "table_hospital", " table_department" and "table_doctor" are shown in Table XVII, Table XVIII and Table XIX respectively. These values will be compared with the values present in these three tables and on finding a match with any value of among these above mention three tables, the corresponding attribute name will be retrieved. The retrieved attribute name will be again compared with value of "attribute_name" field of the table named "table_entity_attribute_sensing". If the retrieved attribute name matches with any value of "attribute_name" field of table named "table_entity_attribute_sensing", that corresponding row value will be fetched from "table_entity_attribute_sensing" and inserted into "table_semantic" for semantic analysis. If these values do not match with above mentioned three tables, they are simply ignored by the proposed system.

Among these six patterns, only the pattern আসানসোল (asansol) matches with the value of "hos_add" field of the table named "table_hospital", the attribute name i.e. "hos_add" will be retrieved. The retrieved "hos_add" attribute again will be compared with the value of "attribute_name" field of the table named "table_entity_attribute_sensing". If the "hos_add" matches with "attribute_name" field of the table named "table_entity_attribute_sensing", then the row value will be fetched and will be inserted into "table_semantic" for semantic analysis. Those patterns which match with table named "table_word", some of them are ambiguous and some are unambiguous.

Patterns খারাপ (kʰarap), চিকিৎসা (tʃikitsa) and হাসপাতাল (ɦaspatal) have single "synset_id" means these patterns are unambiguous and their corresponding "possible_attribute" field's value will be retrieved. These retrieved "possible_attribute" will be compared with value of "attribute_name" field of the table named "table_entity_attribute_sensing". If the retrieved attribute name matches with any value of "attribute_name" field of table named "table_entity_attribute_sensing", then corresponding row value will be fetched from "table_entity_attribute_sensing". The row which is fetched from "table_entity_attribute_sensing", will be inserted into "table_semantic" for semantic analysis. Patterns খারাপ (kʰarap) and চিকিৎসা (tʃikitsa) contain NULL value at their "possible_attribute" field that why these patterns will be ignored by the proposed system. Pattern হাসপাতাল (ɦaspatal) contains "hos_name" at their "possible_attribute" field. The value "hos_name" is compared with "attribute_nane" field value of the table named "table_entity_attribute_sensing", then the row value is fetched and inserted into the "table_semantic" for semantic analysis. But patterns মাথা (mat^{h}a) and মাথা খারাপ (mat^{h}a k^{h}arap) contains more than one synset_ids means these patterns are ambiguous. Now the proposed system applies the LESK algorithm on patterns মাথা (mat^{h}a) and মাথা খারাপ (mat^{h}a k^{h}arap) to extract the correct sense.

The pattern মাথা(mat^{h}a) contains more than one sense in the modified Bengali WordNet DB. Here the proposed system uses retrieved "gloss_concept_defination" and "example_sentence" field value from the table named "table_synset" that has been given below.

Sense 1: gloss_concept_defination –কোন পশুর শরীরের উপরের অংশ যেখানে চোখ,নাক, কানযুক্ত থাকে (the upper part of body which contains eyes, nose, ears etc.)

example_sentence –মাথার চিকিৎসা মনোরোগবিদ্যা বিভাগ আছে সেই হাসপাতালে করানো ভালো (It is better to treat mental diseases in a hospital which has psychiatric department.)

Sense 2: gloss_concept_defination –কোন পশু যে দল, সংস্থা অথবা দেশ নিন্ত্রয়ণ করে

matha – head (head of family, head of a state, leader of a nation, head of a group of humans or animals who controls others.).

example_sentence –গ্রামের মাথাদের কথাই শেষ কথা (The words of the head of a village is final). Here head of a village is human being.

The user given query will be compared with the "gloss_concept_defination" as well as "example_sentence" field value of the pattern মাথা. The sense 1 contains maximum number of overlapping word(s). These overlapping words are মাথা, চিকিৎসা and হাসপাতাল. But the sense 2 does contain only one overlapping word i.e. মাথা. So the sense 1 will be closest sense of the pattern মাথা and the value of "possible_attribute" field of that row of the "table_synset" is to be selected. The retrieved possible attribute name is again compared with value of "attribute_name" field of the table named "table_entity_attribute_sensing". If the retrieved attribute name matches with any value of "attribute_name" field of the table named "table_entity_attribute_sensing", then corresponding row value will be fetched from the "table_entity_attribute_sensing". The row which will be fetched from "table_entity_attribute_sensing", and that will be inserted into "table_semantic" for semantic analysis.

The pattern মাথা খারাপ (mentally challenged) contains more than one sense in the modified Bengali WordNet DB. Similarly, the proposed system will retrieve "gloss_concept_defination" and "example_sentence" field value from the table named "table_synset" that has been given below.

Sense 1: gloss_concept_defination –কোন জীব যে উন্মাদ বা মানসিক অসুস্থ (a mentally challenged living being.)

example_sentence –মাথা খারাপের রোগীকে মানসিকরোগের হাসপাতালের ডাক্তার দ্বারা চিকিৎসা করানো দরকার (mentally challenged people should be treated by psychiatric doctor)

Sense 2: gloss_concept_defination –কোন অনিশ্চতার সমন্ধে উদ্বেগ (to be unusually anxious over a trivial matter.)

example_sentence –তুচ্ছ ব্যপার নিয়ে তুমি মাথা খারাপ করোনা (Do not be anxious over such a small matter.)

The user given query will be compared with "gloss_concept_defination" as well as "example_sentence" field value of the pattern মাথা খারাপ. The sense 1 contains maximum number of overlapping word(s). The overlapping word is চিকিৎসা. But the sense 2 does not contain any overlapping word(s). So the sense 1 will be closest sense of the pattern মাথা খারাপ and the value of "possible_attribute" field of that row of the "table_synset" is to be selected. The retrieved possible attribute name will be compared again with value of "attribute_name" field of the table named "table_entity_attribute_sensing". If the retrieved attribute name matches with any value of "attribute_name" field of table named "table_entity_attribute_sensing", then corresponding row value will be fetched from "table_entity_attribute_sensing". The row which will be fetched from "table_entity_attribute_sensing", and that will be inserted into "table_semantic" for semantic analysis. Our proposed system has disambiguated two Bengali ambiguous words মাথা and মাথাখারাপ asমনোরোগ (mental disease) for current context. Therefore two same entries occur for id 4 and 6 in Table VII. The instance of "table_semantic" for above mention user query has been given in Table VII.

TABLE VII. SENSE EXTRACTION OF PATTERN (S) USING LESK ALGORITHM AND SEMANTIC ANALYSIS

id	entity	attribute_name	primary_key	foreign_key	candidate_key	value
1	table_hospital	hos_add	hos_id			আসানসোল(asansol)(propoer Noun)
2	table_hospital	hos_name	hos_id			
3	table_department	dept_name	dept_id	hos_id		মনোরোগ(mnor[2]og)(mental disease)
4	table_doctor	doc_specialist	doc_id	hos_id	dept_id	মনোরোগ(mnor[2]og)(mental disease)
5	table_department	dept_name	dept_id	hos_id		মনোরোগ(mnor[2]og)(mental disease)
6	table_doctor	doc_specialist	doc_id	hos_id	dept_id	মনোরোগ(mnor[2]og)(mental disease)

Component 7: SQL formation

The proposed system will generate SQL from “table_semantic”. The proposed system will consider only one entry if more than one rows have the same value in entity, “attribute_name”, “primary_key”, “foreign_key”, “candidate_key” and “value” fields. Id 4 and 6 contains same values in“entity”, “attribute_name”, “primary_key”, “foreign_key”, “candidate_key” and “value fields”. One entry will be deleted between id 4 and 6 from the “table_semantic”. Hence the refinement instance of the table named “table_semantic” has been given in Table VIII.

Some predefine postulates have been taken for SQL formation from one of our research article has been described in [6]. The desired result retrieving query in SQL can be given by SELECT attribute 1, attribute 2, attribute 3… attribute n FROM entity 1 (table 1), entity 2 (table 2), entity 3 (table 3)… entity n (table n) WHERE condition 1 and condition 2 and condition 3… and condition n. There are few clauses are fixed any retrieving query. These are SELECT, FROM and WHERE. The proposed system has to determine the correct possible attribute(s), entities and condition(s). Predefine postulates have been described below as well as conditional flowchart has been given in Fig. 3.

TABLE VIII. REFINEMENT TABLE OF SEMANTIC TABLE

id	entity	attribute_name	primary_key	foreign_key	candidate_key	value
1	table_hospital	hos_add	hos_id			আসানসোল(asansol)(propoer Noun)
2	table_hospital	hos_name	hos_id			
3	table_department	dept_name	dept_id	hos_id		মনোরোগ(mnor[2]og)(mental disease)
4	table_doctor	doc_specialist	doc_id	hos_id	dept_id	মনোরোগ(mnor[2]og)(mental disease)

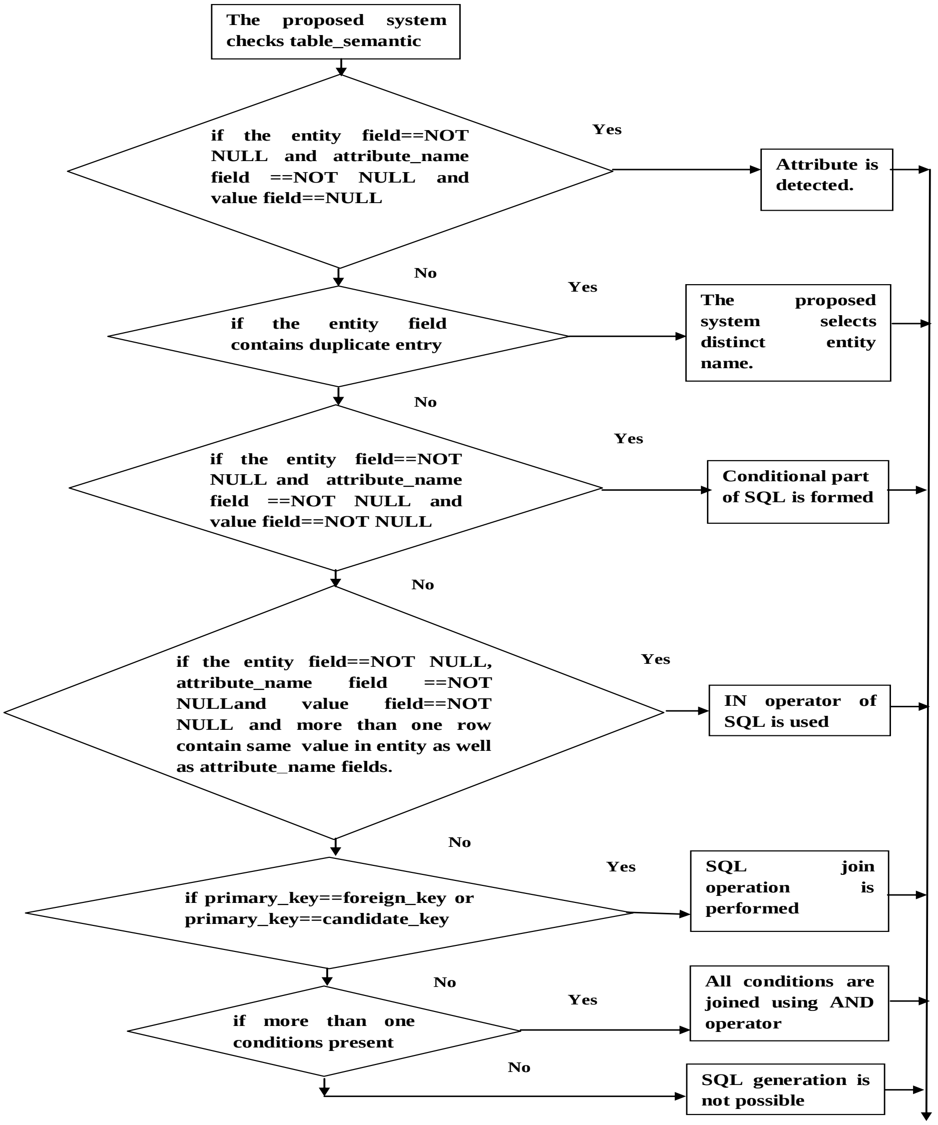

The proposed system checks table_semantic
if the entity field==NOT NULL and attribute_name field ==NOT NULL and value field==NULL
Yes
Attribute is detected.
No
if the entity field contains duplicate entry
Yes
The proposed system selects distinct entity name.
No
if the entity field==NOT NULL and attribute_name field ==NOT NULL and value field==NOT NULL
Yes
Conditional part of SQL is formed
No
if the entity field==NOT NULL, attribute_name field ==NOT NULLand value field==NOT NULL and more than one row contain same value in entity as well as attribute_name fields.
Yes
IN operator of SQL is used
No
if primary_key==foreign_key or primary_key==candidate_key
Yes
SQL join operation is performed
No
if more than one conditions present
Yes
All conditions are joined using AND operator
No
SQL generation is not possible

Fig. 3. Conditional Flowchart for SQL Rules Generation.

Postulate i: The proposed system predicts attribute if the “entity” field, “attribute_name” field and “value” field of the table named “table_semantic” contain NOT NULL, NOT NULL, NULL value respectively, then whatever attribute present in the “attribute_name” field is treated as attribute. In case such type of condition does not occur, then the proposed system considers all attributes i.e. denoted by “*”. Finally this attribute is appended by “.” operator with their corresponding entity field value of that row.

Postulate ii: The system predicts entity name from entity field value(s) of the table named “table_semantic”. For duplicate entry the system selects distinct entity field value.

Postulate iii: The proposed system predicts condition if the “entity” field, “attribute_name” field and “value” field of the table named “table_semantic” does not contain NULL, NULL, NULL value respectively, then whatever attribute present in the “attribute_name” field will be treated as condition. This attribute will be appended by “.” operator with their corresponding entity field value followed by “=” symbol and value of the field value of that row in the semantic table.

Postulate iv: The proposed system determines IN clause in condition if the “entity” field, “attribute_name” field and value field of the table named “table_semantic” does not contain NULL, NULL, NULL value respectively. More than one row in the table named “table_semantic”, the “entity” field as well as corresponding “attribute_name” field contains same value means that particular attribute of that entity has a list of values. Then the attribute will be appended by “.” operator with their corresponding “entity” field value followed by IN clause and a list of values will be placed within opening and closing parenthesis separated by “,”.

Postulate v: The “primary_key” field value of one entity matches with “foreign_key” or “candidate_key” of other entity that means joining occurs.It is a property of relational database. The proposed system appends matched “primary_key” field value by “.” with corresponding “entity” field value followed by “=” symbol and matched “foreign_key” or “candidate_key” value append by “.” with corresponding “entity” field value.

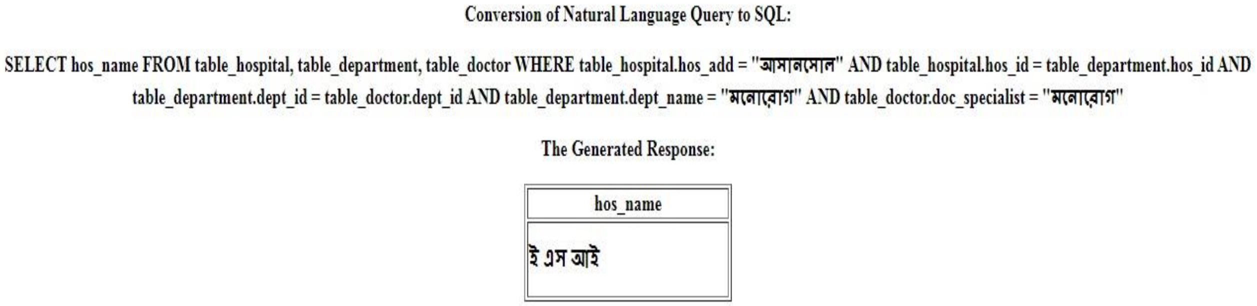
Postulate vi: The proposed system combines all conditions using AND operator if more than one conditions are present.

The proposed system converts the user given Bengali query i.e. আসানসোলে মাথা খারাপের চিকিৎসার জন্য হাসপাতাল কোথায় আছে? into SQL has been given below.

SELECT hos_name FROM table_hospital, table_department, table_doctor WHERE table_hospital.hos_add= 'আসানসোল' AND table_hospital.hos_id = table_department.hos_id AND table_department.dept_id = table_doctor.dept_id AND table_department.dept_name = 'মনোরোগ' AND table_doctor.doc_specialist = 'মনোরোগ'.

Component 8: SQL executed by the proposed system.

The proposed system will execute the SQL that will be generated from user given query i.e. given in Bengali language. After execution of the SQL, the expected result will be generated. The response of the user query1 has been given in Fig. 4.



Conversion of Natural Language Query to SQL:

```
SELECT hos_name FROM table_hospital, table_department, table_doctor WHERE table_hospital.hos_add = "আসানসোল" AND table_hospital.hos_id = table_department.hos_id AND
table_department.dept_id = table_doctor.dept_id AND table_department.dept_name = "মনোরোগ" AND table_doctor.doc_specialist = "মনোরোগ"
```

The Generated Response:

hos_name
ই এস আই

Fig. 4. Response of the Query 1 in Bengali.

VI. Modified Bengali WordNet and Health Information Databases

The data repository of XBLQPS is made up of a modified Bengali WordNet database and a health information database. The modified Bengali WordNet database has been developed with help of Bengali WordNet database as referred in [3]. Table IX, Table XI, Table XII and Table XIII has been kept same as the Bengali WordNet whereas, Table X has been modified by introducing two fields named meaning and “possible_attribute” by the authors. The Health information database contains other tables named “table_inflectional_part”,“table_entity_attribute_sensing”,“table_semantic”,“table_hospital”,“table_department” and “table_doctor” which have been introduced and developed by the authors. The table named “table_semantic”, “table_hospital”, “table_department” and “table_doctor” have been taken from our research article referred in [6]. The proposed system thus has modified the original WordNet database of [3] and shall now be termed as “Modified Bengali WordNet DB”. Each word in the Modified Bengali WordNet DB contains its parts of speech definition introduced manually as according to the largest probability of its use. We have not referred to any standard Bengali POS Tagger but to our own prepared POS definitions in Modified Bengali WordNet DB. The Modified Bengali WordNet DB and Health information database consist of a set of tables as follows:

Structure of table “table_word”

The table named “table_word” consists of two fields’ “word_id” and “word_name”. The field “word_id” is primary key field whereas “word_name” field contains Bengali word.Data types and sizes of these two fields are int(10) and varchar(100) respectively. The instance of the table “table_word” has been given in Table IX.

TABLE IX. STRUCTURE OF TABLE “TABLE_WORD”

word_id	**word_name**
7	বিভাগ(bib^{ɦ}aɡ) (department)
16	অপথ্যালমোলজি （ɔpt^{h}ɛlmolodʒi）(ophthalmology)
23	মাথাখারাপ(mat^{h}ak^{h}arap) (mental disease)

Structure of table “table_synset”

The table named “table_synset” consists of “synset_id”, “parts_of_speech”, “gloss_concept_defination”, “example_sentence”, “meaning” and “possible_attribute”. The “synset_id” is the primary key field of this table. The “parts_of_speech” field contains possible parts of speech value of the Bengali word which is stored in table “table_word”. The “gloss_concept_defination”, “example_sentence” and “meaning” field contain definition, example of sentence and meaning of above mention Bengali word respectively. The “possible_attribute” field contains likelihood relationship of a particular Bengali word with medical domain. If a “possible_attribute” field does not contain any value that means it has NULL value.Data types and sizes of these six fields are int(10), varchar(50), varchar(150), varchar(200), varchar(300) and varchar(150) respectively. The instance of the table “table_ synset” has been given in Table X.

TABLE X. STRUCTURE OF TABLE “TABLE_SYNSET”

synset_id	**parts _of_ speech**	**gloss _concept_ definition**	**example_sentence**	**meaning**	**possible_attribute**
2	বিশেষ্য(biʃeʃo)(noun)	একটি প্রতিষ্ঠানের বিশেষ অংশ	রবি রয় অস্থিচিকিৎসা বিভাগে রডাক্তার		dept_name
5	বিশেষ্য(biʃeʃo)(noun)	চক্ষুরোগ সমন্ধীয় চিকিৎসা	চক্ষুরোগে আক্রান্ত ব্যক্তি রাত্রে চশমা ব্যবহার করেন	চক্ষুবিজ্ঞান	dept_name,doc_specialist
10	ক্রিয়া(kir[2]ia) (verb)	কোন অনিশ্চতার সমন্ধে উদ্বেগ	তুচ্ছ ব্যপার নিয়ে তুমি মাথা খারাপ করোনা	বিরক্ত করা	

Structure of table “table_sense”

The table named “table_sense” has two fields. These two fields are “word_id” and “synset_id”. The primary key field consists of both fields. This table is used to map a “word_id” to its corresponding “synset_id”. Data types and sizes of these two fields are int(10) and int(10) respectively. The instance of the table “table_ sense” has been given in Table XI.

TABLE XI. STRUCTURE OF TABLE “TABLE_SENSE”

word_id	**synset_id**
7	2
16	5
23	10

Structure of table “table_hypernym”

The table named “table_hypernym” consists of two fields. These two fields are “synset_id” and “hypernym_id". The “synset_id” field stores the synsetid value of a Bengali word which is stores in the table “table_word” whereas “hypernym_id” field value is nothing but a “synset_id” of a another word. These two words have hypernym and hyponym relationship.Data types and sizes of these two fields are int(10) and int(10) respectively. The instance of the table “table_hypernym” has been given in Table XII.

TABLE XII. STRUCTURE OF TABLE “TABLE_HYPERNYM”

synset_id	**hypernym_id**
7	3
8	3

Structure of table “table_tree”

The table named “table_tree” consists of two fields. These two fields are “hypernym_id” and “parent_id”. The “hypernym_id” is “synset_id” of a Bengali word. The “parent_id” is parent word id of a hypernym word. This table maintains the hierarchical relationship between hypernym and hyponym word. The parent_id field value contains zero that represents root word of the tree. Data types and sizes of these two fields are int(10) and int(10) respectively. The instance of the table “table_tree” has been given in Table XIII.

TABLE XIII. STRUCTURE OF TABLE “TABLE_TREE”

hypernym_id	**parent_id**
3	0

Structure of table “table_entity_attribute_sensing”

The table named “table_entity_attribute_sensing” consists of “id”, “entity”, “attribute_name”, “primary_key”, “foreign_key” and “candidate_key”. The “id” field is the primary key field of this table. The “entity” field contains name of participating entities in medical domain. The “attribute_name” field stores attribute name, “primary_key” field contains primary key, “foreign_key” contains foreign key and “candidate_key” stores candidate key value of corresponding entity. The sometimes foreign key and candidate key may not exist, that time these key fields contain NULL value.The “table_entity_attribute_sensing” is an independent table. This tablehas been used to map entity from attribute, or primary key, foreign key, candidate key from entity.Data types and sizes of these six fields are int(10), varchar(50), varchar(50), varchar(50), varchar(50) and varchar(50) respectively. The instance of the table “table_entity_attribute_sensing” has been given in Table XIV.

TABLE XIV. STRUCTURE OF "TABLE_ENTITY_ATTRIBUTE_SENSING"

id	entity	attribute_name	primary_key	foreign_key	candidate_key
1	table_hospital	hos_name	hos_id		
2	table_hospital	hos_add	hos_id		
3	table_hospital	hos_district	hos_id		

Structure of table "table_inflectional_part"

The table named "table_inflectional_part" consists of "id" and "inflectional_part". The id field is primary key field of this table. The "inflectional_part" field stores inflectional part of the Bengali word. Data types and sizes of these two fields are int(10) and varchar(50) respectively. The instance of the table "table_inflectional_part" has been given in Table XV.

TABLE XV. STRUCTURE OF "TABLE_INFLECTIONAL_PART"

id	inflectional_part
1	টি(t̪i)
2	ে(e)
3	ের(er[2])

Structure of table "table_semantic"

The table named "table_semantic" includes id, entity, "attribute_name", "primary_key", "foreign_key", "candidate_key" and "value" fields. The "id" field is the primary key of this table. The entity name, corresponding attribute name, primary key, foreign key, candidate key and value are held by fields named"entity", "attribute_name", "primary_key", "foreign_key", "candidate_key" and "value" respectively. The "table_semantic" is an independent table. This table has been used to construct the SQL.Data types and sizes of these seven fields are int(10), varchar(50), varchar(50), varchar(50), varchar(50), varchar(50) and varchar(100) respectively. The instance of the table "table_semantic" has been given in Table XVI.

TABLE XVI. STRUCTURE OF "TABLE_SEMANTIC"

id	entity	attribute_name	primary_key	foreign_key	candidate_key	value
1	table_hospital	hos_add	hos_id			আসানসোল(asansol) (propoer Noun)
2	table_hospital	hos_name	hos_id			
3	table_department	dept_name	dept_id	hos_id		মনোরোগ(mnor[2]og)(mental disease)

Structure of table "table_hospital"

The table named "table_hospital" includes "hos_id", "hos_name", "hos_add", "hos_district", and "hos_state" fields. The "hos_id" field is the primary key of this table. This "hos_id" field used to identify uniquely each entity instance of the table. The hospital name, hospital address, district name and state name where hospital is situated, are held by fields named "hos_name", "hos_add", "hos_distrct", and "hos_state" respectively.Data types and sizes of these five fields are int(10), varchar(100), varchar(100), varchar(100) and varchar(100) respectively. The instance of the table named "table_hospital" has been given in Table XVII.

TABLE XVII. STRUCTURE OF "TABLE_HOSPITAL"

hos_id	hos_name	hos_add	hos_district	hos_state
1	মুর্শিদাবাদজেলাহাসপাতাল (mur[2]ʃidabaddʒelaɦaspatal) (Murshidabad District Hospital)	লালগোলা(lalgola) (Lalgola) (proper noun)	মুর্শিদাবাদ (mur[2]ʃidabad) (Murshidabad) (proper noun)	পশ্চিমবঙ্গ(pʃt̪ʃimbɔŋg) (West Bengal) (proper noun)
8	হাওড়াজেলাহাসপাতাল (ɦaoɽadʒelaɦaspatal) (Howrah District Hospital)	আমতা(amta) (Amta) (proper noun)	হাওড়া (ɦaoɽa) (Howrah)(proper noun)	পশ্চিমবঙ্গ(pʃt̪ʃimbɔŋg) (West Bengal) (proper noun)

Structure of table "table_department"

The table named "table_department" includes "dept_id", "dept_name", and "hos_id"fields.Data types and sizes of these three fields are int(10), varchar(100) and int(10) respectively. The instance of the table "table_department" has been given in Table XVIII.

TABLE XVIII.STRUCTURE OF "TABLE_DEPARTMENT"

dept_id	dept_name	hos_id
70	নবজাতক(nbdʒatk) (neonate)	7
170	মনোরোগ(mnor[2]og) (psychiatry)	12

Structure of table "table_doctor"

The table named "table_doctor" includes "doc_id", "doc_name", "doc_qualification", "doc_ specialist", "hos_id" and "dept_id" fields. Data types and sizes of these six fields are int(10), varchar(100), varchar(100), varchar(100), int(10) and int(10) respectively. The instance of the table "table_doctor" has been given in Table XIX.

TABLE XIX. STRUCTURE OF "TABLE_DOCTOR"

doc_id	doc_name	doc_qualification	doc_specialist	hos_id	dept_id
1000	কেয়া গরাই(kea gr[2]ai̯) (Keya Gorai)	এম.বি.বি.এস. (M.B.B.S.)	অপথ্যালমোলজি(ɔptʰɛlmolodʒi) (Ophthalmologist)	1	10
1010	শমীতা দাশগুপ্ত(ʃmita daʃgupta) (Shamita Dasgupta)	এম.ডি. (M.D.)	অস্থি(ɔstʰi) (Orthopaedist)	2	20

VII. TOOLS AND INTERNATIONAL PHONETIC ALPHABET (IPA) NOTATION USED

Software tools like HTML, PHP, MySQL and Avro Bengali keyboard has been used to developed the XBLQPS. The HTML has been used to develop the web pages structure. A sever side scripting language i.e. PHP has been used to provide customize interface as well as process the user request. MySQL is used as back end database to store the data. The IPA notation for Bengali language has been taken from the website https://en.wikipedia.org/wiki/Help:IPA/Bengali.

VIII. TIME COMPLEXITY

The time complexity has been calculated for algorithmic steps of the proposed system.

Step 1: Login into the system and submit the query in Bengali language.

Let, as an example query আসানসোলে মাথা খারাপের চিকিৎসার জন্য হাসপাতাল কোথায় আছে?. The time complexity has been calculated on the proposed algorithm using the above example query. The time has been calculated in each algorithmic step that is given below.

Step 2: Tokenization

Let, there be n number of token(s) in user given query has been given in Fig. 5.

Time taken for tokenization of n token(s) = n unit time.

The number of tokens will be generated from above example query has been given in Table XX.

Here for this query time taken for tokenization=8 unit time because number of token(s) i.e. n=8.

Step 3: Root word extraction

Let, there be p numbers(s) of inflection part in the table “table_inflectional_part”, and also assume that q number(s) of token(s) which are matched with value of “inflectional_part” field.

∴ Time taken= p×q unit time.

Here p=3, q=3.

Therefore, time taken=3×3=9-unit time.

Step 4: POS tagging

Let, there be r number(s) of “word_name” in the “table_word”.

Let, there be r number of “word_id” and “synset_id” in the table “table_sense”.

Let, there be r number(s) of parts of speech in the table “table_synset”.

∴Time taken = (p×r+r×r+r×r) unit time.

Here r=3 for “table_word”.

r=3 for “table_sense”.

r=3 for “table_sysnset”.

∴Time taken= (3×3+3×3+3×3) unit time=27-unit time.

Step 5: Pattern generation of noun phrases

Let, n number(s) of token(s) are participating for pattern generation where the token occurrence orders are to be maintained.

∴Time taken=n+(n-1) +(n-2)+⋯+1-unit time.

=n/2 (n+1)

There are 11 numbers of patterns which are already generated in Table V.

Therefore, time taken = 11 unit time.

Step 6: Sense extraction of pattern(s) using LESK algorithm and semantic analysis.

Time taken for unambiguous pattern={n+(n-1)+(n-2)+⋯+1}(r+r+r) unit time.

=n/2 (n+1)(r+r+r)

Let, there be u number(s) of similar “word_id” in the “table_sense”.

Let, v number(s) of “synset_id” contains “possible_attribute” field value among u number(s) of similar “synset_id”.

Let, x number(s) of values exists “attribute_name” in the table “table_entity_attribute_sesing”.

Let, y number(s) of instance(s) have been inserted in the table “table_semantic”.

Time taken for ambiguous pattern={ n+(n-1)+(n-2)+⋯+1}(r+r×u+u×r+v×x+v×y) unit time.

Time taken for unambiguous patterns =8(3+3+3)=72 unit time.

u=number of similar “word_id” in the “table_sense” =3.

v=3, x=3, y=6.

∴Time taken for ambiguous patterns= 8(3+3×3+3×3+3×3+3×6) unit time=384 unit time.

1	2	3	…	n - 2	n - 1	n

Fig. 5. Token(s) of user given Query.

TABLE XX. TOKEN OF USER GIVEN QUERY.

Number of tokens	1	2	3	4	5	6	7	8
Token of example query	আসানসোলে (asansole)	মাথা (matʰa)	খারাপের (kʰaraper)	চিকিৎসার (tʃikitsar)	জন্য (dʒno)	হাসপাতাল (ɦaspatal)	কোথায় (kotʰae̯)	আছে (atʃʰe)

Step 7: SQL formation

Let, time taken for SQL formation=z unit time.

6 postulates have been applied on 4 rows of “table_semantic” to form SQL.

∴Time taken =(4×6) unit time =24 unit time.

Step 8: SQL executed by the proposed system

Let, time taken for SQL execution=t unit time.

∴Time complexity =f(n,p,q,r,t,u,v,x,y,z)

=n+p×q+(p×r+r×r+r×r)+n+(n-1)+(n-2)+⋯+1+{n+(n-1)+(n-2)+⋯+1}(r+r+r)+{n+(n-1)+(n-2)+⋯+1}(r×u)+r×u+u×r+v×x+v×y+z+t

∴f(n)≅n+n×n+(n×n+n×n+n×n)+n+(n-1)+(n-2)+⋯+1+{n+(n-1)+(n-2)+⋯+1}(n+n+n)+{n+(n-1)+(n-2)+⋯1}n+n×n+n×n+n×n+n×n+n+n

$$= n + n^2 \; + 3n^2 + \frac{n}{2}(n+1) + \left\{\frac{n}{2}(n+1)\right\}(3n) + \left\{\frac{n}{2}(n+1)\right\} n + n^2 + n^2 + n^2 + n^2 + n + n = n + 4n^2 + \frac{n^2}{2} + \frac{n}{2} + \frac{3n^2}{2}(n+1) + \frac{n^2}{2}(n+1) + 4n^2 + n + n$$

$$= n + 8n^2 + \frac{n^2}{2} + \frac{n}{2} + \frac{3n^3}{2} + \frac{3n^2}{2} + \frac{n^3}{2} + \frac{n^2}{2} + n + n$$

$$= O(n^3)$$

Time taken for SQL execution t=11 unit time.

∴Time complexity algorithm.

= (8+9+27+11+72+384+24+11) unit time.

=546 unit time.

IX. Precision, Recall and Accuracy Analysis of the XBLQPS

Example of Precision, recall and accuracy analysis of the proposed XBLQPS.

111 numbers of queries have been submitted into the proposed system by the user. Four different types of queries based on their output have been given as examples. The output has been classified into “expected output” and “select output”. “expected output” and “select output” have been further categorized into positive and negative. The Confusion Matrix [23], [24] for the proposed system has been given in Table XXI.

TABLE XXI. Confusion Matrix

Expected output Vs. Select output	**Select output-Positive**	**Select output-Negative**
Expected output-Positive	True positives (TP)-46	False positives (FP)-16
Expected output-Negative	False negatives (FN)-17	True negatives (TN)-32

Some categorized sample queries has been given below as an example.

TP- আসানসোলে মাথা খারাপের চিকিৎসার জন্য হাসপাতাল কোথায় আছে?

TN- বাকুড়া জেলাহাসপাতালে অ্যাম্বুলেন্সের ব্যবস্থা আছে?

FP- সুমিত রায়ের শিক্ষাগত যোগ্যতা কি?

FN-বাকুড়া হাসপাতাল

Precision = TP/(TP+FP)=46/(46+16)=46/62=0.74

Our system has a precision of 0.74 - in other words, when it predicts a query is fetched, it is correct 74% of the time.

Recall = TP/(TP+FN)=46/(46+17)=46/63=0.73

Our system has a recall of 0.73 - in other words, it correctly identifies 73% of all fetched queries.

Accuracy=(TP+TN)/(TP+FP+FN+TN)=(46+32)/(46+16+17+32)=78/111=0.70

Our system has an accuracy of 0.70 - in other words, it gives 70% correct predictions of fetched queries.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)=(2*0.73*0.74)/(0.73+0.74)=1.08/1.47=0.73

Our system has F1 score is 0.73 – in other words, overall measure of our system’s accuracy is 73%.

X. Future Work

The XBLQPS uses a simplified version of the LESK algorithm. It depends on the maximum number of the common word(s) between word gloss and the current context of the word. As the gloss may be limited for a domain in the proposed Bengali WordNet, it may happen that the proposed system sometimes fails to disambiguate an ambiguous word. Our proposed system uses a manually annotated parts of speech method for the detection of parts of speech of a word which is based on the largest probability of its use. The proposed system needs to be improved by using a modified version of the LESK algorithm and automated POS tagging method for better handling natural language queries with ambiguous words and this can be marked as a future work of this proposed system.

XI. Conclusion

The XBLQPS uses a simplified version of the LESK algorithm. It depends on the maximum number of the common word(s) between word gloss and the current context of the word. As the gloss may be limited for a domain in the proposed Bengali WordNet, it may happen that the proposed system sometimes fails to disambiguate an ambiguous word. Our proposed system uses a manually annotated parts of speech method for the detection of parts of speech of a word based on the largest probability of its use. In the future, the proposed system needs to be improved by using a modified version of the LESK algorithm and automatic POS tagging method for better efficiency wherein the gloss shall be enhanced to accommodate several synonymous applications of a word in that particular domain. Moreover, we may enhance the proposed system to work with unstructured databases. Retrieving data from a relational database management system from a Bengali language query containing the ambiguous word is a challenging task. The proposed XBLQPS will be able to handle a domain-specific Bengali language query containing an ambiguous word which will be helpful even for a naive user. The naïve user can access the database without knowledge of for-mal language i.e. SQL (Structured Query Language). The XBLQPS is an enhancement and advanced form of BLQPS of the work discussed in [6]. The XBLQPS incorporates the handling of ambiguous Bengali words using the LESK algorithm and a modified WordNet in Bengali. The time complexity, precision, recall, accuracy and F1 score have been analyzed for our proposed system. The time complexity of XBLQPS is $O(n^3)$ as compared to the time complexity of BLQPS of [6] which is $O(n^4)$ which shows that the time efficiency of XBLQPS is better than BLQPS.

ACKNOWLEDGMENT

This study was conducted at the National Institute of Technology (NIT), Durgapur, Research Project Lab, which is part of the Department of Computer Science and Engineering. The authors express their gratitude to the Department of Computer Science and Engineering, NIT, Durgapur, India, for providing educational resources for this research.

REFERENCES

[1] H.K. Azad, and A. Deepak, “A new approach for query expansion using Wikipedia and WordNet,” Elsevier, Vol. 492, pp. 147-163, 2019.

[2] N.S. Dash, P. Bhattacharyya, and J.D. Pawar, “The WordNet in Indian Languages,” Springer, pp. 243-260, 2017.

[3] F. Faruqe, and M. Khan, “Bwn-a software platform for developing bengali wordnet,” Innovations and Advances in Computer Sciences and Engineering., Springer, pp. 337-342, 2010.

[4] S. Madonsela, “African Wordnet as a tool to identify semantic relatedness and semantic similarity,” South African Journal of African Languages, Taylor & Francis, vol. 39, no. 2, pp. 185-190, 2019.

[5] O. Dongsuk, S. Kwon, K. Kim, Y. Ko, “Word sense disambiguation based on word similarity calculation using word vector representation from a knowledge-based graph,” Proceedings of the 27th international conference on computational linguistics, pp. 1-12, 2018.

[6] K.P Mandal, P. Mukherjee, A. Chattopadhyay, B. Chakraborty, “A novel Bengali Language Query Processing System (BLQPS) in medical domain,” Intelligent Decision Technologies, IOS Press, vol. 13, no. 2, pp. 177-192, 2019.

[7] G. Huilin, G. Tong, W. Fan, M. Chao, “Bidirectional Attention for SQL Generation,” 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 676-682, 2019.

[8] P. Sharma, N.J.E. Joshi, Technology, and A.S. Research, “Knowledge-Based Method for Word Sense Disambiguation by Using Hindi WordNet,” Engineering, Technology and Applied Science Research, vol. 9, no. 2, pp. 3985-3989, 2019.

[9] A. Sau, T.A. Amin, N. Barman, A. R. Pal, “Word Sense Disambiguation in Bengali Using Sense Induction,” 2019 International Conference on Applied Machine Learning (ICAML), IEEE, pp. 170-174, 2019.

[10] M.M. Rahman, S.A. Khan, and K.A. Hasan. “Word Sense Disambiguation by Context Detection,” 2019 4th International Conference on Electrical Information and Communication Technology (EICT), IEEE, pp. 1-6, 2019.

[11] H. Walia, A. Rana, and V. Kansal. “Case based interpretation model for word sense disambiguation in Gurmukhi,” 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, pp. 359-264, 2019.

[12] M.S. Kaysar, M. A. B. Khaled, M. Hasan, M. I. Khan, “Word sense disambiguation of Bengali words using FP-growth algorithm,” 2019 international conference on electrical, computer and communication engineering (ECCE), IEEE, pp. 1-5, 2019.

[13] G. Jain, D. Lobiyal, “Word Sense Disambiguation of Hindi Text using Fuzzified Semantic Relations and Fuzzy Hindi WordNet,” 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, pp. 494-497, 2019.

[14] N.S. Dash, “Back to Basics: A Road to Return to Nominal Base through Lemmatization,” Proceedings of Abstracts of the 36th International Conference of the Linguistic Society of India (ICOLSI-36), pp. 1-34, 2014.

[15] S. Basuki, A. S. Kholimi, A. E. Minarno, F. D. S. Sumadi, M. R. A. Effendy, “Word Sense Disambiguation (WSD) for Indonesian Homograph Word Meaning Determination by LESK Algorithm Application,” 2019 12th International Conference on Information & Communication Technology and System (ICTS), IEEE, pp. 8-15, 2019.

[16] Y. Heo, S. Kang, J.J.I.A. Seo, “Hybrid sense classification method for large-scale word sense disambiguation,”IEEE, vol. 8, pp. 27247-27256, 2020.

[17] C. Lachichi, C. Bendiaf, L. Berkani, A. Guessoum, “An Arabic WordNet enrichment approach using machine translation and external linguistic resources,” 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP), IEEE, pp. 1-6, 2018.

[18] Q.-P. Nguyen, A.-D. Vo, J.-C. Shin, P. Tran, C.-Y. Ock, “Korean-vietnamese neural machine translation system with korean morphological analysis and word sense disambiguation,”, IEEE, vol. 7, pp. 32602-32616, 2019.

[19] A. Jabbar, S. Iqbal, A. Akhunzada, Q. Abbas, “An improved Urdu stemming algorithm for text mining based on multi-step hybrid approach,” Journal of Experimental & Theoretical Artificial Intelligence, Taylor & Francis, vol. 30, no. 5, pp. 703-723, 2018.

[20] Q. Zhou, Y. Meng, “combination of Semantic Relatedness with Supervised Method for Word Sense Disambiguation,” 2019 International Conference on Asian Language Processing (IALP), IEEE, pp. 142-147, 2019.

[21] M. Biswas, M.M. Hoque, “Development of a Bangla Sense Annotated Corpus for Word Sense Disambiguation,” 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), IEEE, pp. 1-6, 2019.

[22] A. Haque, M.M.J.I. Hoque, “Bangla word sense disambiguation system using dictionary based approach,” pp. 1-6, 2016.

[23] A. Jakka, J.J.I.J.I.T.E.E. Vakula Rani, “Performance Evaluation of Machine Learning Models for Diabetes Prediction,” International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 11, pp. 1976-1980, 2019.

[24] A. Santra, J. Christy, “Genetic algorithm and confusion matrix for document clustering,” International Journal of Computer Science Issues, vol. 9, no. 1, pp. 322, 2012.

[25] K. JanakiRaman, K. Meenakshi, “Automatic Text Summarization of Article (NEWS) Using Lexical Chains and WordNet,” International Journal of Advanced Science and Technology, Vol. 29, no. 4, pp. 3242-3258, 2020.

[26] A. Alkhatlan, J. Kalita, A. Alhaddad, “Word Sense Disambiguation for Arabic Exploiting Arabic WordNet and Word Embedding,” The 4th International Conference on Arabic Computational Linguistics (ACLing 2018), Elsevier, Vol. 142 pp. 50-60, 2018.

[27] A. H. Aliwy, A. R. Abbas, “Improvement WSD Dictionary Using Annotated Corpus and Testing it with Simplified Lesk Algorithm,” Fifth International conference on Computer Science and Information Technology, pp. 89-97, 2015.

AUTHORS' PROFILE

Kailash Pati Mandal his obtained Bachelor of Technology in Computer Science and Engineering from Bengal College of Engineering and Technology, Durgapur, India in 2006. In 2011, he received Master's degree in Computer Science and Engineering from the Jadavpur University, Kolkata, India. He is currently a part-time Researcher in Computer Science and Engineering in the field of Natural Language Processing at the National Institute of Technology (NIT), Durgapur, India.

Prasenjit Mukherjee obtained his PhD in Computer Science and Engineering from National Institute of Technology (NIT), Durgapur, India under Visvesvaraya PhD scheme Program, Ministry of Electronics and Information Technology (MeitY), Govt. of India. His research interests include machine learning, deep learning, natural language processing, knowledge engineering, and database management systems. He has more than 15 international publications.

Atanu Chattopadhyay received the M.Sc degree in Applied Mathematics and Computing from Indian Institute of Technology (Indian School of Mines), Dhanbad, India. He is presently a fulltime Lecturer in the Department of BBA(H) & BCA (H), Deshabandhu Mahavidyalaya, Chittaranjan, West Bengal, India. His research interest includes Mathematical Analysis and Natural Language Processing.

Baisakhi Chakraborty obtained her PhD in Computer Science and Engineering from National Institute of Technology, Durgapur, India in 2011. Her research interests include knowledge systems, knowledge engineering and management, database systems, data mining, natural language processing, and software engineering. She has several researchers under her direction. She has more than 30 international publications. She has a decade of industry experience and 17 years of academic experience.

MSDAR: Multi-Stage Dynamic Architecture Intrusion Detection System

Ahmed M. ElShafee[1]
Ahram Canadian University
Cairo, Egypt

Marianne A. Azer[2]
National Telecommunication Institute
Nile University, Cairo, Egypt

***Abstract*—Ad hoc networks have been through extensive research in the last decade. Even with their desirable characteristics, major issues related to their security need to be considered. Various security solutions have been proposed to reduce the risks of malicious actions. They mainly focus on key management, authentication, secure localization, and aggregation techniques. These techniques have been proposed to secure wireless communications but they can only deal with external threats. Therefore, they are considered the first line of defense. Intrusion detection systems are always required to safeguard ad hoc networks as such threats cannot be completely avoided. In this paper, we present a comprehensive survey on intrusion detection systems in ad hoc networks. The intrusion detection systems and components and taxonomy as well as different implementations and types of IDSs are studied and categorized. In addition, we provide a comparison between different Intrusion Detection Systems' architectures. We also propose a Multi Stage Dynamic Architecture intrusion detection system (MSDAR), designed with a multi-stage detection approach making use of both signature-based and anomaly detection benefits. Our proposed intrusion detection system MSDAR is featured by its dynamic architecture as it can be deployed in the network using the Distributed Hierarchical Architecture. The viability and performance of the proposed system MSDAR are tested against the Distributed Denial of Service Attacks through simulations. Advanced performance parameters were used to evaluate the proposed scheme MSDAR. Experimental results have shown that the performance of MSDAR improves by using multiple stages of different detection mechanisms. In addition, based on simulations, the Detection Rate increases when the sensitivity level increases.**

Keywords—Ad hoc networks; attacks; DDoS; intrusion detection; security

I. Introduction

Emerging technologies have contributed in revolutionizing our daily life. To mention a few, Artificial Intelligence (AI) [1], Blockchain, cryptocurrencies [2], Internet of Things (IoT) [3], cloud computing, and wireless technology. Wireless technology is critical to today's communications [4], and essential to developing technologies within the next years. Wireless communications are almost based on ad hoc or special purpose connections. Mobile Ad hoc Networks (MANETs) are key players in the future of wireless communication [5]. They consist of distributed nodes without any predetermined infrastructure [6]. The lightweight mobile devices have the capabilities of sensing and processing received information [7]. The devices have a limited transmission range that needs intermediate nodes to reach other far nodes. Due to their special features, ad hoc networks are susceptible to a wide range of attacks [8], exterior and interior threats and misbehaving modes [9]. Some of these attacks are initiated to deprive legitimate users of network services. Other attacks have the objective of gaining unauthorized access to network resources [10].

MANETs have many different challenges regarding designing security solutions due to their vulnerability to eavesdropping, lack of trusted management, limited computation capabilities, and power sources which increase their vulnerability to Denial of Service (DoS) attacks and also can become incapable of running heavy security algorithms. Due to the open, self-organized, infrastructure-less environment of MANETs, there is a chance that trusted nodes to be hijacked. Therefore, any security solution should be designed to defend the network against both insider and outsider attacks. In MANETs, insider attacks are more problematic and difficult to overcome. Security solutions for ad hoc networks are considered to be one of the most active and attractive research areas. Researchers mainly focus on key management [11], authentication, secure localization, and aggregation techniques to secure wireless communications [12]. The current security solutions can only deal with external threats, and therefore they can be considered the first line of defense. However, insider attackers that already exist within the first perimeter of defense can penetrate the whole network and cause severe damage. Therefore, Intrusion Detection Systems are considered the second line of defense as they come into action after the intrusion has already occurred [13]. There are two types of Intrusion Detection Systems (IDSs), signature based detection IDS, and anomaly-based detection IDS [14]. Signature-based IDSs (misuse detection) require a knowledge base containing the behavioral patterns of different attacks. When the IDS detects a certain pattern that refers to an attack, it alerts the network's users against this specific attack. The main disadvantage in such implementation is that only known attacks are caught and reported. This may surge the percentage of false negatives. On the other hand, anomaly detection, IDSs (behavior-based detection) are not designed to catch threats using their signature or pattern. They are developed to learn the normal behavior patterns of both users and network applications to discover and report any altered patterns. In anomaly-based IDSs, new and unknown attacks can be detected and reported whenever they occur. However, any abnormal benign behaviors will be caught and reported as new threats. This may increase the percentage of false positives.

This research introduces a newly developed trust-based IDS for wireless ad hoc networks. The proposed Multi Stage Dynamic Architecture Intrusion Detection System (MSDAR) takes into consideration multistage detection mechanisms to increase its capability to detect different types of intrusions. The first and third stages are based on anomaly detection, while the second is signature-based detection. In the third stage, an additional parameter is used, it is called the sensitivity level.

The contributions of this paper are as follows.

1) The Intrusion Detection Systems components and taxonomy as well as different implementations and types of IDSs are studied and categorized. The study's objective is to understand the algorithms and design parameters and their impact on performance and functionality.

2) A comparison between different Intrusion Detection Systems' architectures from the points of view of complication, precision, scalability, and possibility of failure is provided.

3) The taxonomy of IDSs' architectures, detection algorithm, and additional design parameters is presented.

4) Our proposed intrusion detection system architecture and the operational algorithm are explained in detail. The proposed MSDAR is designed with a multi-stage detection approach making use of both signature-based and anomaly detection benefits. Simulation analysis methodology, simulation parameters, simulation metrics used for performance evaluation, and simulation results are also presented and discussed in detail.

The remainder of this paper is organized as follows: In Section II, we give an overview of the related work done for securing ad hoc networks using Intrusion Detection Systems. Section III gives an insight into our implementation. Section IV presents the simulation results and evaluation of our proposed scheme MSDAR. Section V is focused on discussing the results and mentioning the limitations Of MSDAR. Finally, Section VI concludes the paper and presents future directions are presented.

II. Related Work

Intrusions are any kind of unauthorized or unapproved activities within the network. Intrusion Detection Systems are schemes and tools, used to discover, assess and report intrusions that may compromise the network. IDSs should continuously adapt and improve, to be able to discover new attacks and attack strategies. Many factors have motivated the development of IDSs. First, the presence of security flaws and vulnerabilities in a complex system makes it susceptible to malicious intrusions. Second, is the inefficiency of most of the prevention techniques that were designed and implemented to prevent possible attacks. Third, the exposure to insider attacks is expressed to be much more harmful than outsider attacks, even in most secure systems. Finally, newly emerged attacks need considerably advanced security solutions. This makes IDSs an attractive and important research area. For this research, different implementations and types of IDSs are studied and categorized in this section. The study's objective is to understand the algorithms and design parameters and their impact on performance and functionality, to overcome any unexpected flaws in our new proposed technique MSDAR. This section is organized as follows. Structural components and building blocks are introduced in Section A. Section B gives an insight into IDSs' architecture taxonomy. Supplementary design parameters and their taxonomy are introduced in section C.

A. Intrusion Detection Systems' Components

Because of their common goals, most of IDSs share the same structural patterns [15]. Data collecting and formatting, analysis and detection, and reaction mechanism units are the three primary parts of any IDS. The main components of the intrusion detection systems are depicted in Fig. 1. Various data types from different sources are collected, formatted and sorted at the data collection and formatting unit and then delivered to the analysis and detection unit. Collected data is analyzed and processed and then compared to the normal system behavior in anomaly-based IDS, or the signature of known attacks in signature-based IDS, or finally the well-defined specifications of a program or protocol in specification-based IDS [16]. After an action is detected as malicious, it is reported to the response mechanism. The response mechanism is defined according to the designed response policy. Different responses can only be categorized into two groups; passive responses and active responses. The passive response is done by simply notifying the authorized entity of the identified malicious action or intrusions detected. On the other hand, an active response is any form of action aiming to mitigate the threat or expected damage resulting from an intrusion or attack. This can be done by terminating network connections for certain periods or blocking IP addresses/Physical addresses linked to the attack. The response policy should also illustrate the response period as it can be either permanent or temporary. IDSs with active response mechanisms can also be aliased as Intrusion Prevention Systems (IPSs).

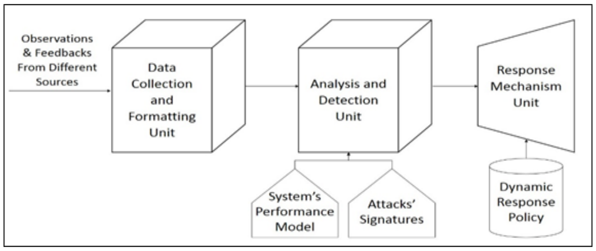

Observations & Feedbacks From Different Sources
Data Collection and Formatting Unit
Analysis and Detection Unit
Response Mechanism Unit
System's Performance Model
Attacks' Signatures
Dynamic Response Policy

Fig. 1. The Intrusion Detection System's Building Blocks.

B. Intrusion Detection Systems Architectures Taxonomy

Intrusion detection systems can be deployed in the network using different architectures [17]. These architectures can be classified into two broad categories; Standalone, and collaborative, as shown in Fig. 2. Early IDSs were implemented as stand-alone systems having only local monitors and analysis units at each node. Local monitors and analysis units serve only their host nodes by detecting abnormal events according to the predefined detection policy. The response against any action is addressed and limited to the node's level with no extra extension. Stand-alone IDSs are not immune against distributed attacks and they can't be reliable for detecting malicious events occurring simultaneously at different locations inside the network. Therefore, there is a need for Collaborative IDSs. In collaborative architectures, an IDS enforces cooperation between monitors to provide a considerably more scalable and accurate model than stand-alone IDS. Collaborative IDS are also classified according to the communication model between both monitor units and analysis units as depicted in Fig. 2. Collaborative IDSs classification includes four subcategories, centralized, decentralized, distributed, and finally, our newly proposed architecture that is illustrated in this paper; hierarchically-distributed. Centralized IDSs, depend on one single centralized analysis unit in addition to several distributed monitoring units at each node or entity in the network. Two main disadvantages of such IDSs are the scalability limitations and the Single Point of Failure (SPoF). This is because the single analysis unit can handle only a limited number of monitoring units and it can be an easy target for direct attacks to disable the entire functionality of the intrusion detection system. Decentralized IDSs make use of multiple analysis units distributed in different locations within the network. Each analysis unit is responsible for accumulating, aggregating, and analyzing data from different monitoring units. Finally, a head analysis unit on top of all other analysis units receives this information, to make nondiscriminatory decisions regarding network entities and events. Such architecture supports scalability and overcomes the bottleneck congestion presented in centralized IDSs. In Distributed IDSs, each entity in the network is equipped with a monitor unit and an analysis unit. Each node shares its information with its peers in a completely distributed model. Collected data are organized and analyzed among all nodes. In Distributed IDS architecture, both congestion and SPoF disadvantages are avoided. However, an extra processing requirement is added to each node within the network. This additional requirement may consume extra processing and power capabilities during intrusion detection activities which in turn minimize the nodes' capabilities required to process normal flow. Therefore, a new architecture is proposed to overcome the disadvantages mentioned above. It is based on both Distributed and Hierarchical Architecture (DHA-IDS) as shown in Fig. 3.

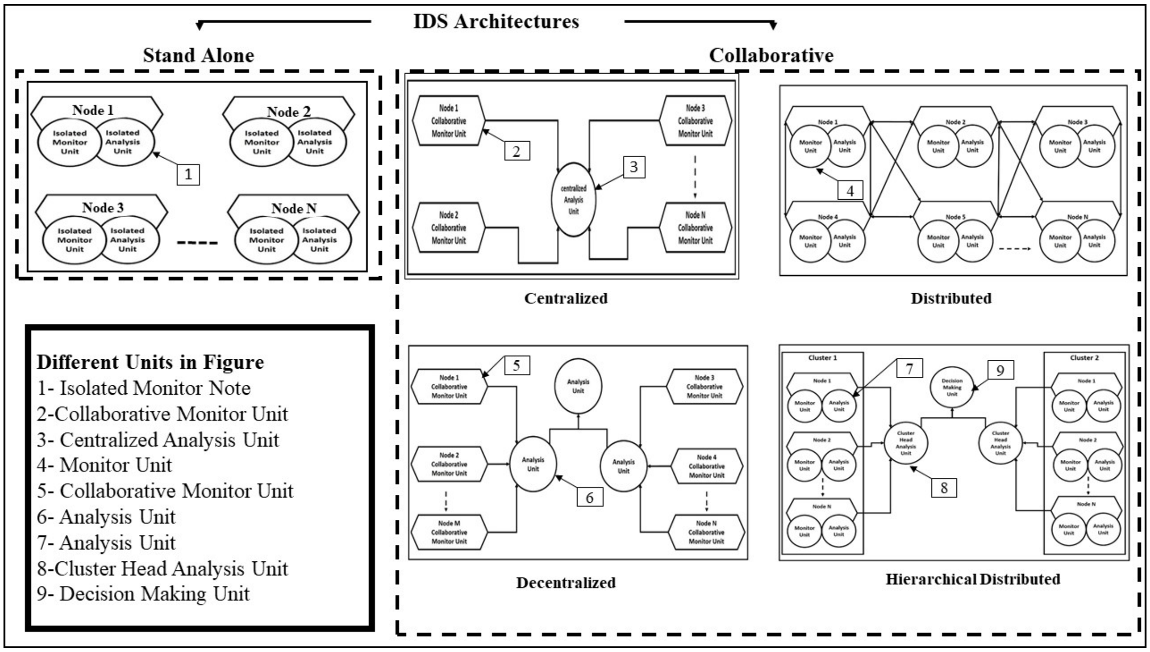

IDS Architectures
Stand Alone
Collaborative
Node 1
Isolated Monitor Unit
Isolated Analysis Unit
Node 2
Node 3
Node N
1
Node 1 Collaborative Monitor Unit
Node 2 Collaborative Monitor Unit
Node 3 Collaborative Monitor Unit
Node N Collaborative Monitor Unit
centralized Analysis Unit
2
3
Centralized
Monitor Unit
Analysis Unit
Node 4
Node 5
4
Distributed
Different Units in Figure
1- Isolated Monitor Note
2-Collaborative Monitor Unit
3- Centralized Analysis Unit
4- Monitor Unit
5- Collaborative Monitor Unit
6- Analysis Unit
7- Analysis Unit
8-Cluster Head Analysis Unit
9- Decision Making Unit
Node M Collaborative Monitor Unit
5
6
Decentralized
Cluster 1
Cluster 2
Decision Making Unit
Cluster Head Analysis Unit
7
8
9
Hierarchical Distributed

Fig. 2. Extended Intrusion Detection Systems Architectures Taxonomy.

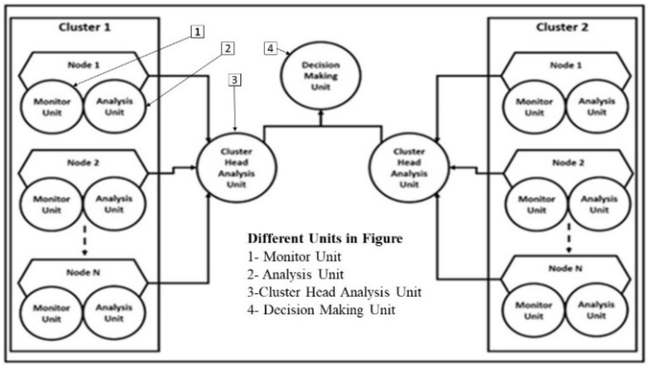


Fig. 3. Proposed Distributed Hierarchical Intrusion Detection System (DHA-IDS) Model.

In DHA-IDS, each node has its own monitor and analysis units. Each node is responsible for monitoring and analyzing only data collected by itself, then it forwards the analyzed data about different network activities to the cluster head analysis unit directly above it. Each cluster head analysis unit is responsible for collecting data from the nodes within its cluster. It then runs the second phase of processing and analyzing to correlate all collected information. At the end, each cluster head analysis unit forwards the correlated information to the response unit which is responsible of deciding a proper action related to detected intrusions. If an attack is directed to the response mechanism unit, one of the cluster head analysis units will become responsible of replacing the response mechanism unit. Similarly, in case the attack is extended to the cluster heads, each node will depend on its information and make its own decision regarding any suspected action. Therefore, this proposed DHA-IDS can perform in the worst attack conditions and it can degrade its architectural level from Distributed Hierarchical to Standalone, in order to retain the system's self-robustness. Furthermore, the new proposed architecture overcomes many disadvantages of different architectures mentioned above like; bottleneck congestion, SPoF, processing and power overheads. Table I compares between the different architectures presented in this section and the proposed DHA-IDS. They are compared based on complexity, accuracy, scalability, and risk of failure.

TABLE I. COMPARISON BETWEEN DIFFERENT IDSS' ARCHITECTURES USING THE FOLLOWING CRITERIA: COMPLICATION, PRECISION, SCALABILITY, POSSIBILITY OF FAILURE ON THREE LEVELS, LOW (L), MEDIUM (M), HIGH (H)

Architecture	**Complexity**			**Precision**			**Scalability**			**Possibility of Failure**		
Classification	**L**	**M**	**H**	**L**	**M**	**H**	**L**	**M**	**H**	**L**	**M**	**H**
Stand alone	√			√					√	√		
Centralized	√					√	√					√
Decentralized		√		√					√	√		
Distributed			√		√				√		√	
Distributed-Hierarchical		√				√			√	√		

C. Intrusion Detection Systems Design Parameters Taxonomy

Different parameters are taken into consideration when a new IDS is designed. These parameters influence the IDS performance. Some of these parameters are presented in the following sections.

1) Source of data: According to the source of data, IDSs can be classified as Host-based, Network-based, and Hybrid. In stand-alone architectures, data is collected and analyzed locally from each node independently. Another approach is collaborative security, which is accomplished by multiple correlated sources. Data can be collected either locally and independently, or globally from each node (host-based) in the network then the collected data can be correlated and analyzed in a holistic form [18]. In such IDSs, monitoring units are deployed locally in the host to detect host-targeted attacks. Examples of such attacks are the ones aiming to exhaust the hosts' resources or gain unauthorized access to systems' components and data. Data can also be collected from network traffic (network-based) instead of nodes' local data [19]. Monitoring units are deployed in firewalls, or routers to capture all network packets. This information can help to detect different threats and possible attacks and spot abnormal activities in the network. Finally, some IDSs depend on their design and implementation on both sources of data, host-based, and network-based. This type of IDS is described as (hybrid-sourced). Hybrid-sourced IDSs can detect various types of attacks targeting any of the host or network components.

2) Scheduling of analysis: The intrusion detection process can trail different schedules (real-time, offline). In real-time analysis [20], data is collected, then immediately correlated and analyzed. Instantly, an appropriate decision is taken regarding the detected behavior. On the other hand, offline analysis [21] is performed after all nodes forward their collected data to the analysis unit. While the data is being analyzed, the nodes pursue their normal operation. Whenever they receive a decision concerning the network activities, they act accordingly.

3) Initiation: Nodes in IDSs can voluntarily participate in the intrusion detection process like in proactive systems. Monitoring units collect data that is automatically forwarded to analysis units. On the contrary, in driven systems, nodes wait for a direct request to send their own data regarding any activity in the network. Also, the passive nodes don't request neighboring nodes' data. They only receive data passing by their perimeter passively.

4) Types of shared data: Collaborative IDSs depend on sharing data among the network elements. There are three types of data: Raw, partially processed, and fully processed. Raw data is collected by nodes that are not equipped with any analysis units. Data is then forwarded to other nodes with higher processing capabilities for analysis. Environmental data and behavior logs are examples of raw data. In case of partially processed data, nodes are more powerful, so they can be used to minimize the traffic overhead due to forwarding every single piece of raw data at each node in the network. Also, IDSs make use of partially processed data to minimize the processing capabilities required by each node in the IDS's higher levels. Finally, the existence of malicious or abnormal activity in the network is determined by the fully processed data. Therefore, confirmed intrusions, attacks, decisions, and alerts regarding a node can be considered fully processed data. Fig. 4 depicts the taxonomy of IDSs' Architectures, detection algorithms, and additional design parameters. Table II summarizes the classification and the comparison between some IDSs that have been proposed in the literature with respect to different parameters such as Communication architecture, and detection algorithms. This is in addition to other design parameters such as data source, shared data type, scheduling of analysis, and response mechanism point of view.

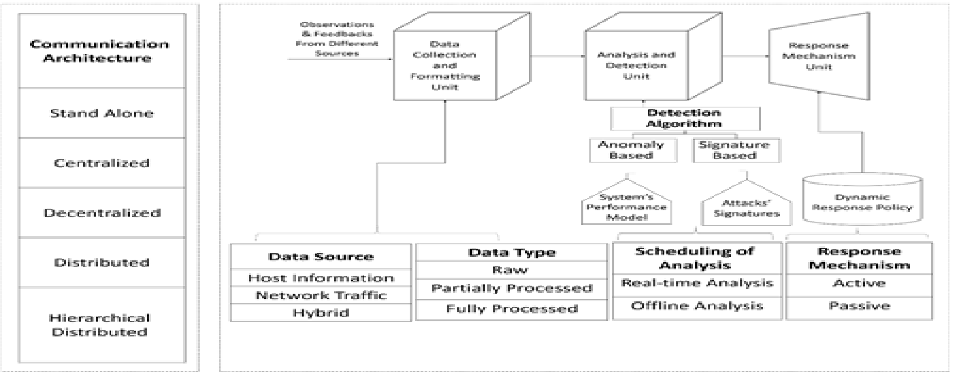


Fig. 4. Taxonomy of IDSs' Architectures, Detection Algorithm and Additional Design Parameters.

TABLE II. COMPARISON BETWEEN INTRUSION DETECTION SYSTEMS BASED ON THEIR DESIGN PARAMETERS

ntrusion Detection System			**DIDS [22]**	**CRIM [23]**	**DIDMA [24]**	**GRIDS [25]**	**HIDE [26]**	**EMERALD [27]**	**INDRA [28]**	**LARSID [29]**	**DOMINO]30]**	**SNORT [31]**	**WORMI -NATOR [32]**	**QUI CK SAND [33]**	**PEREZ [34]**	**LIDS [35]**
Communication Architecture		**Centralized**	√	√	√											
		Decentralized				√	√	√							√	√
		Distributed							√	√	√	√	√	√		
Detection Algorithm		**Signature Based**		√	√	√		√			√	√		√	√	√
		Anomaly Based	√				√	√	√	√			√			
Additional Design Parameters	**Data Source**	**Host Based**		√	√	√		√		√				√		√
		Network Traffic						√				√				
		Hybrid	√				√		√		√		√		√	
	Data Type	**Raw**		√						√						
		Partially Processed	√		√	√					√				√	
		Fully Processed					√	√	√			√	√	√		√
	Analysis Schedule	**Real-time**				√	√		√	√		√				
		Offline	√	√	√			√			√		√	√	√	√
	Response Mechanism	**Active**							√		√	√				
		Passive	√	√	√	√	√	√		√			√	√	√	√

III. Proposed System Architecture

This section presents our proposed system architecture and operational algorithm. The proposed MSDAR is designed with a multi-stage detection approach. The first stage is implemented using anomaly detection with a classifier mechanism. The system in this stage has a statistical prediction of most successive events in the network. The analysis unit compares the current event to the pre-predicted event, if they match; then the system is considered to be operating in its normal state. If the current event doesn't match any of the pre-predicted events, then it will be considered an anomaly. The second stage is implemented using a signature-based detection mechanism. At this phase, the analysis unit compares the current event - detected as an anomaly in the first stage - to the predefined attacks behaviors' and signatures' profiles. Therefore, an anomaly detected in the first stage is considered as the audit data for the second stage. The third stage and the following ones are implemented using anomaly detection with a classifier mechanism that has an additional parameter taken into consideration. This parameter is the Sensitivity level of upcoming comparisons. Sensitivity Level SL is incremented each time the system needs to make further investigations regarding the group of events suspected to be an intrusion and correspondingly to minimize the false positive intrusion percentage. The state diagram of the proposed system MSDAR is depicted in Fig. 5. MSDAR follows the standardized structure of known collaborative IDSs. Data collection and formatting units are the first components of its structure. It is designed to have various data sources distributed through the network, and at each host within the network perimeter. Monitor unit and mini-analysis unit are implemented at different data sources. Collected and formatted data are forwarded to the main analysis and detection unit where the analysis processes follow the scheme shown in the flow chart depicted in Fig. 6. Finally, the response unit has the role of propagating a proper response related to any detected intrusion. The response is decided according to the designed response policy. The active response is considered against any intrusive action.

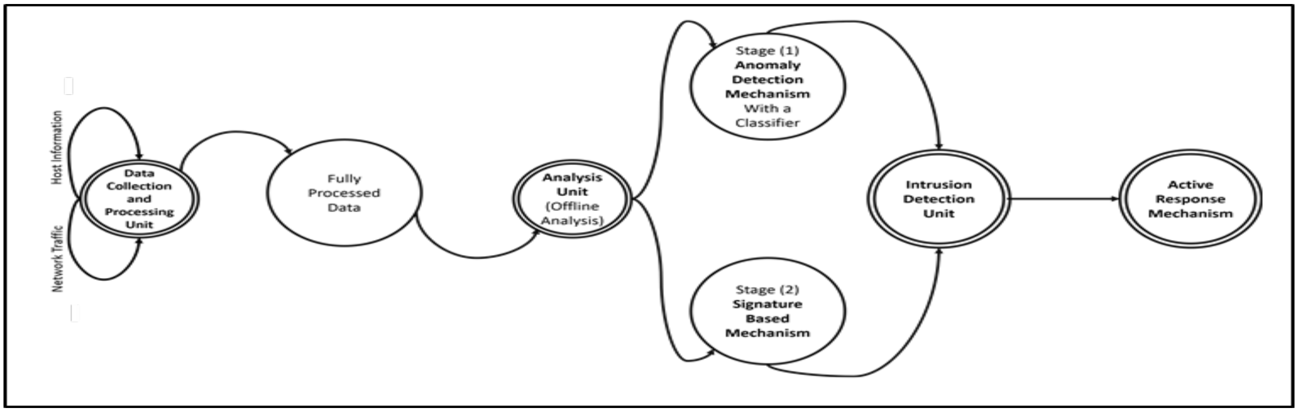

Host Information
Network Traffic
Data Collection and Processing Unit
Fully Processed Data
Analysis Unit (Offline Analysis)
Stage (1) Anomaly Detection Mechanism With a Classifier
Stage (2) Signature Based Mechanism
Intrusion Detection Unit
Active Response Mechanism

Fig. 5. State Diagram of Multi Stage-Dynamic Architecture-IDS (MSDAR).

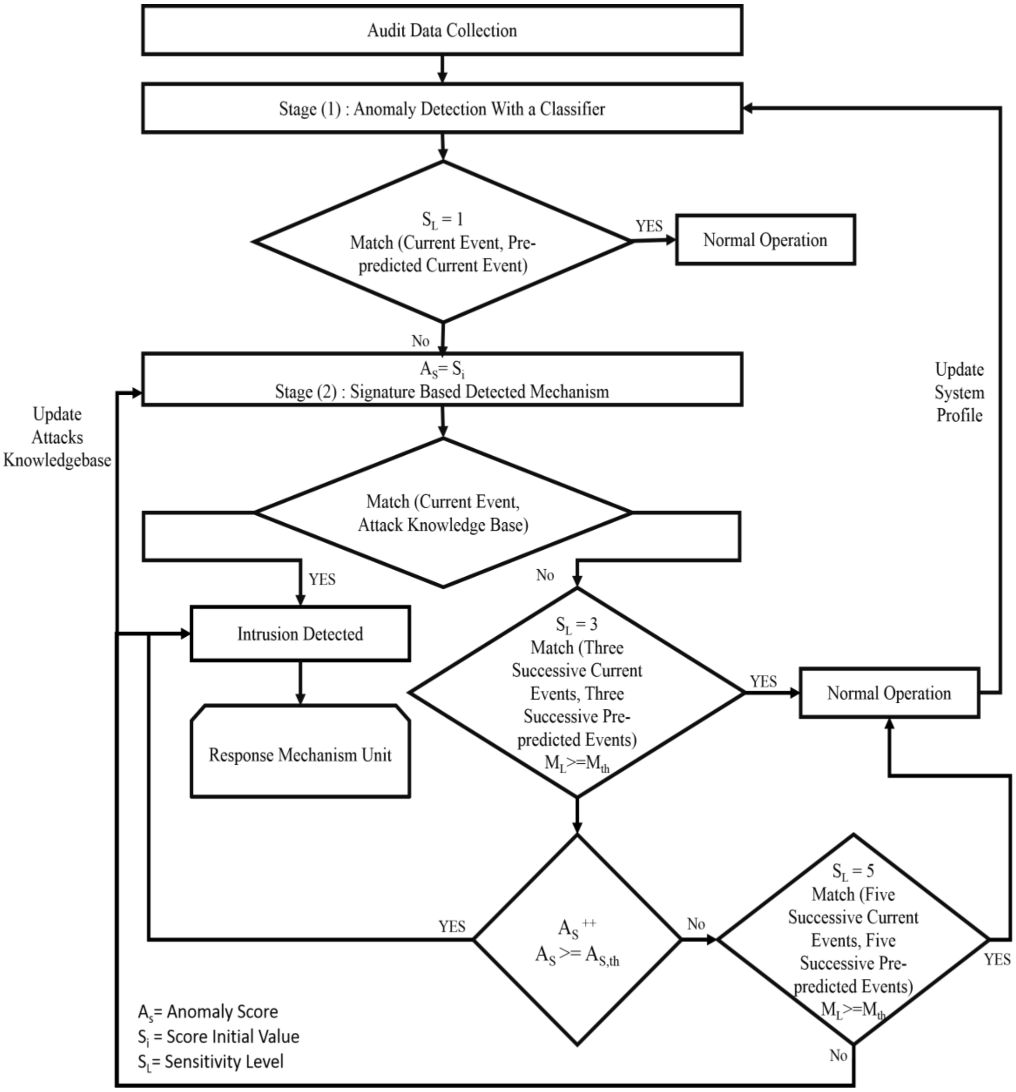

Audit Data Collection
Stage (1) : Anomaly Detection With a Classifier
$S_L = 1$
Match (Current Event, Pre-predicted Current Event)
YES
Normal Operation
No
$A_S = S_i$
Stage (2) : Signature Based Detected Mechanism
Update Attacks Knowledgebase
Update System Profile
Match (Current Event, Attack Knowledge Base)
YES
No
Intrusion Detected
Response Mechanism Unit
$S_L = 3$
Match (Three Successive Current Events, Three Successive Pre-predicted Events)
$M_L >= M_{th}$
YES
Normal Operation
YES
A_S^{++}
$A_S >= A_{S,th}$
No
$S_L = 5$
Match (Five Successive Current Events, Five Successive Pre-predicted Events)
$M_L >= M_{th}$
YES
No
A_s = Anomaly Score
S_i = Score Initial Value
S_L = Sensitivity Level

Fig. 6. FULL Operational Flow Chart of Multi Stage-Dynamic Architecture-IDS (MSDAR).

IV. MSDAR Simulation Results

This section describes the technique used to evaluate MSDAR performance, using the OMNET++ Simulator. The simulated network consists of 80 nodes acting as routers, two workstations acting as data sources, and one application server acting as the victim. All nodes are assigned static IP addresses to enable the possibility of tracking routing tables at each node.

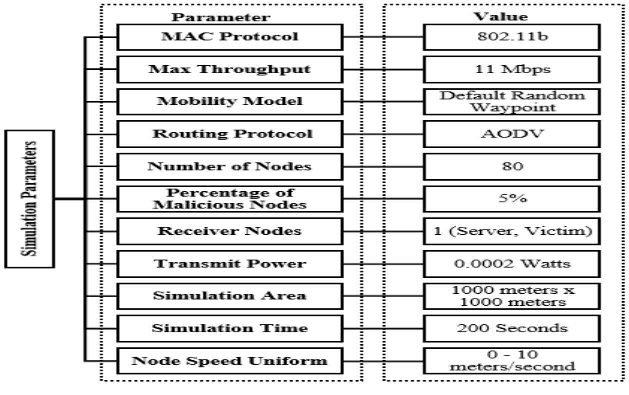
For the simulation, AODV is used as the routing protocol. Fig. 7 summarizes the simulation parameters. For the attack scenario, a certain percentage of nodes are manipulated to act as malicious attackers. The monitored systems' behavioral statistics are then gathered. The following statistics are included: total packets transferred, total packets received, total packets deleted, total packets changed, latency, total connections to the victim node, and average throughput… etc).

A. Simulation and Analysis Methodology

Each event that scores greater than the predefined threshold is marked as an intrusion. Subsequently, a proper action using the MSDAR response mechanism is initiated. In the final stage of the simulation process, overall network performance evaluation is presented in graphs to validate MSDAR's ability to detect the existence of intrusive actions. The following tables present the parameters used for the simulation, testing scenario, and collect statistical data respectively. To assess the efficacy of our suggested system, it is important to measure its ability to distinguish between intrusive and non-intrusive activities, with a minimum number of false alarms. In our evaluation, we adopted the approach in [36]. The metrics used are defined in Table III [36]. The previously mentioned singular metrics were used to form new performance measures. These performance measures are introduced in Table IV [37]. Some researchers consider DR and TPR as the same measure: the proportion of intrusive events that were identified as attacks to all other normal events [37].

To have a fair comparison between IDSs performance, the authors in [36] proposed a metric called Capability of Intrusion Detection (CID) based on some of the metrics mentioned in Table IV, according to (1).

$$CID = -B(1-\beta)\log(PPV) - B(\beta)\log(1-NPV) - (1-B)(1-\alpha)\log(NPV) - (1-B)(\alpha)\log(1-PPV) \quad (1)$$



Parameter	Value
MAC Protocol	802.11b
Max Throughput	11 Mbps
Mobility Model	Default Random Waypoint
Routing Protocol	AODV
Number of Nodes	80
Percentage of Malicious Nodes	5%
Receiver Nodes	1 (Server, Victim)
Transmit Power	0.0002 Watts
Simulation Area	1000 meters x 1000 meters
Simulation Time	200 Seconds
Node Speed Uniform	0 - 10 meters/second

Simulation Parameters

Fig. 7. Simulation Parameters.

TABLE III. Metrics Defined for IDS's Performance Evaluation [36]

Met ric	Meaning	Explanation
FP	False Positive.	The probability of having an alert while no intrusion occurs.
TP	True Positive.	The likelihood of receiving an alarm during an incursion.
FN	False Negative.	The likelihood of not receiving an alarm when an incursion occurs.
TN	True Negative.	The likelihood of not receiving an alert if no incursion happens.
PPV	Positive Predictive value.	The likelihood that an intrusion results in an alert.
NPV	Negative Predictive Value.	The likelihood of no inclusion results in no alert.
B	Base Rate.	The likelihood of an intrusion in the audit data gathered.

TABLE IV. Advanced Performance Measures for IDS's Evaluation [37]

Performance Parameter	Definition	Equation	Value Range
Classification Rate (CR)	The ratio between accurately classified events and the total number of events.	$\frac{TP+TN}{TP+TN+FP+FN}$	CR > 0 CR < 1
Detection Rate (DR)	The proportion of properly identified attacks to the total number of intrusive occurrences.	$\frac{TP}{TP+FN}$	DR > 0 DR < 1
False Positive Rate (FPR) (α)	The proportion of non-intrusive events detected as attacks to the total number of non-intrusive occurrences.	$\frac{FP}{FP+TN}$	FPR > 0 FPR < 1
True Positive Rate (TPR) (1-FNR) (1-β)	The proportion of intrusive events detected as attacks to the total number of regular occurrences.	$\frac{TP}{FP+TN}$	TPR > 0 TPR < 1

B. Simulation Results

One of the most difficult tasks is data gathering [38]. In our simulations, we used the dataset DARPA 2000 Lincoln Laboratory Scenario (LLDDoS) 2.0 which is provided by MIT [39]. It consists of a DDoS attack run by five attackers. A number of simulation sessions are used to carry out this assault scenario. Over time, these sessions were organized into 5 attack stages. MSDAR has been simulated and tested against LLDDOS 2.0.2. The following graphs have been deduced from the simulations. Fig. 8 illustrates the point-to-point throughput during the five time phases of the attack. It shows five peaks at each attack incidence. Fig. 9 illustrates the number of connections directed at the victim node. It can be noticed that the number of connections is exponentially increasing with time. Fig. 10 and Fig. 11 demonstrate the average throughput of the network and received throughput at the victim node respectively. The Receiver Operating Characteristic (ROC)

Curve [40] is the detection rate as a function of the false positive rate and the corresponding calculated CID curve as a function of false positive rate for the different stages (sensitivity levels) of MSDAR respectively. Fig. 12 depicts the ROC of our scheme. From Fig. 12, it can be concluded that the Detection Rate increases when the sensitivity level increases. For αavg = 0.5 we achieved average DR = 0.48, 0.68 and 0.9 at SL = 1, 3 and 5 respectively. The ROC curve is not useful in determining the optimal operation point of MSDAR. On the contrary, the optimal operation point for each stage is declared by the CID curve shown in Fig. 12. Table V shows the maximum CID Levels corresponding to different parameters. It can be deduced that the performance of MSDAR improves by using multiple stages of different detection mechanisms.

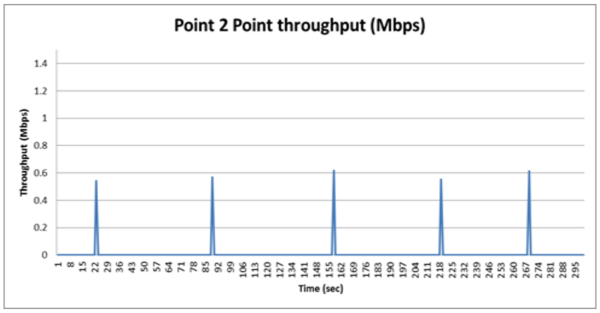


Fig. 8. Point to Point Throughput of the Simulated Network as measured by MSDAR

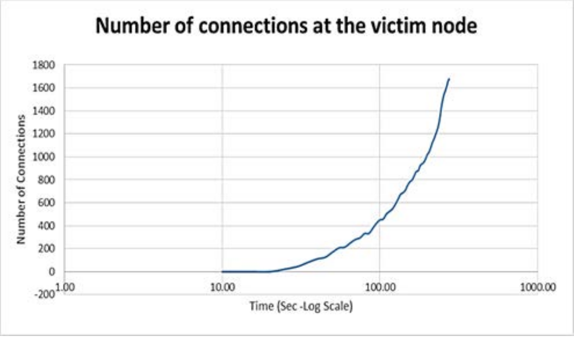


Fig. 9. Number of Connections at the Victim Node as measured by MSDAR.

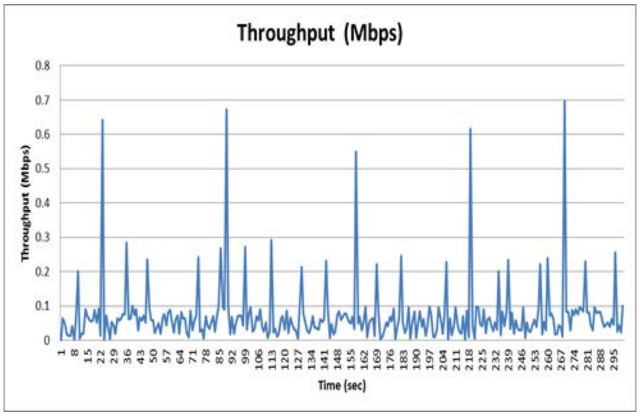


Fig. 10. Average Throughput of the Simulated Network as measured by MSDAR.

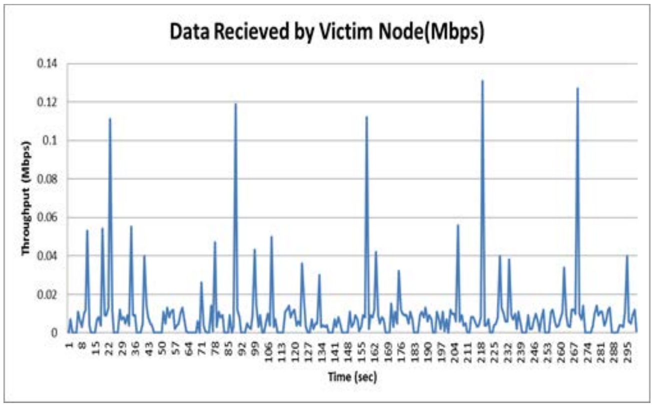
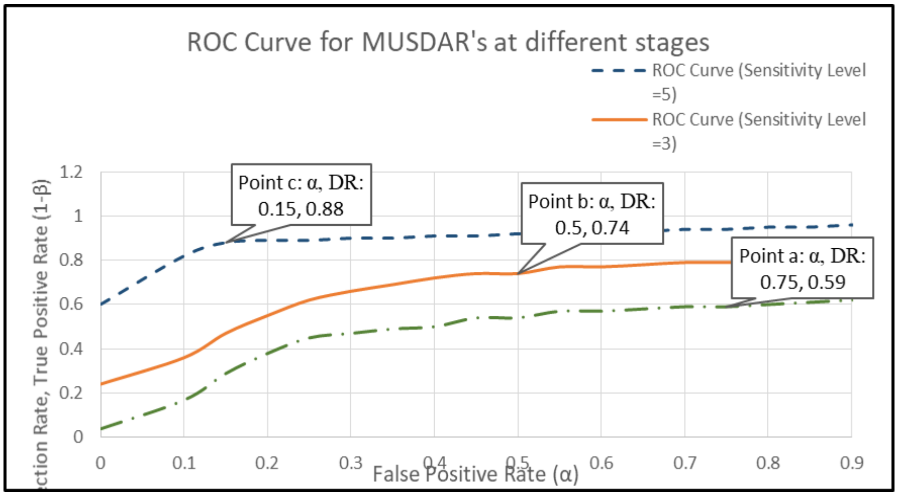


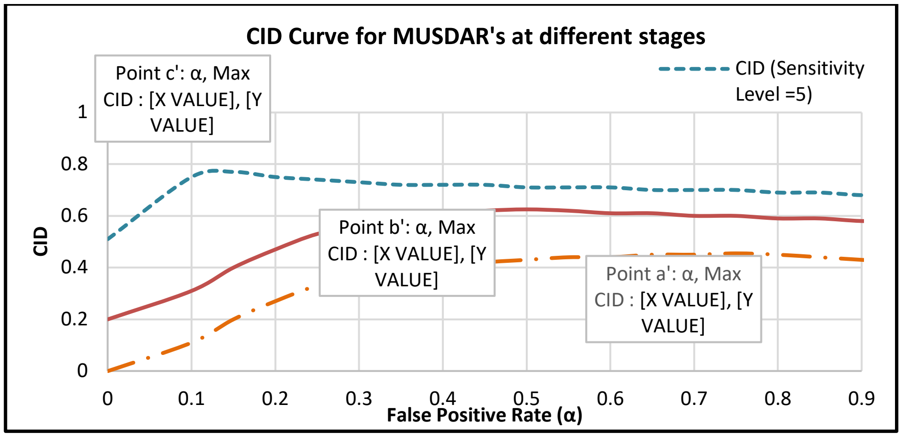
Fig. 11. Data Received by the Victim Node as measured by MSDAR.

TABLE V. MAXIMUM CID LEVELS CORRESPONDING TO DIFFERENT PARAMETERS

Point	α	SL	Maximum CID Level
a'	0.75	1	0.455
b'	0.5	3	0.625
c'	0.15	5	0.77



(a)



(b)

Fig. 12. MSDAR ROC Curve and its Corresponding Capability of Intrusion Detection (CID).

V. Discussion and Limitations

Our newly developed Multi Stage Dynamic Architecture (MSDAR) was explained in detail from the points of view architecture, sequence diagram, and flow chart. It was tested against DDoS attacks through simulations. Each event that scores greater than the predefined threshold is marked as an intrusion. Subsequently, a proper action using the MSDAR response mechanism is initiated. An important factor in evaluating the effectiveness of our proposed system was its ability to distinguish between intrusive and non-intrusive activities, with minimum false alarms. Results have shown that by increasing the IDS sensitivity level, the detection rate increases. The optimal operation point for each stage is declared by the CID curve. This research can be extended by using the statistical test (t-test/p-test/ANOVA to compare and benchmark our method with others. Machine learning models have been used for intrusion detection for over a decade [40]. We plan to use machine learning in one of the stages of our multistage Intrusion detection system MSDAR [41]. We will also use the model from [42] to assess the effectiveness of our suggested approach using machine learning.

VI. Conclusion and Future Work

Despite the various applications of Mobile Ad Hoc Networks, security challenges need to be addressed for both internal and external attacks. Intrusion detection systems are regarded as the second line of security against many types of attacks. Due to MANETs' special characteristics, traditional Intrusion detection systems cannot be used. In this paper, we distinguished between the different approaches used for intrusion detection mechanisms in a structured way. We classified intrusion detection systems with respect to different categories, such as architectures and design parameters. We introduced a standardized building block for intrusion detection systems for MANETs that summarizes different classifications of IDS techniques. In addition, a survey that shows the most popular design parameters used in different IDSs was presented. We proposed a multi-stage intrusion detection system (MSDAR) which is featured by its dynamic architecture as it can be deployed in the network using the Distributed Hierarchical Architecture (DHA-IDS), as it can dynamically change its deployment architecture. Simulations have shown that in case of an attack directed to the response mechanism unit, one of the cluster head analysis units is responsible for replacing the response mechanism unit. Similarly, in case the attack is extended to the cluster heads, each node depends on its data and makes its own decision regarding any suspected action.

Therefore, the proposed MSDAR can perform in the worst attack conditions and it can modify its architectural level from Distributed Hierarchical to Standalone, in order to retain the system's self-robustness. Furthermore, the new proposed architecture is capable of incapacitating many disadvantages of different architectures like; bottleneck congestion, single point of failure, processing, and power overhead. The suggested system's MSDAR effectively lowers false positives, increasing the intrusion detection system's capability and detection rate (CID) are increased by using the multi-stage feature. By measuring the CID level and comparing it to the detection rate, we were able to determine the optimal operation point for each stage in the proposed system. As a result, the total detection rate rises, increasing the network's functional efficiency to a tolerable level. In future work, MSDAR can be tested for different types of attack scenarios.

References

[1] Ahmed I, Jeon G, Piccialli F. From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where. IEEE Transactions on Industrial Informatics. 2022 Jan 27;18(8):5031-42.

[2] A. Nasir A, Shaukat K, Khan KI, Hameed IA, Alam TM, Luo S. What is core and what future holds for blockchain technologies and cryptocurrencies: A bibliometric analysis. IEEE Access. 2020 Dec 23;9:989-1004.

[3] Shaukat K, Alam TM, Hameed IA, Khan WA, Abbas N, Luo S. A review on security challenges in internet of things (IoT). In2021 26th International Conference on Automation and Computing (ICAC) 2021 Sep 2 (pp. 1-6). IEEE.

[4] Shaukat K, Iqbal F, Hameed IA, Hassan MU, Luo S, Hassan R, Younas A, Ali S, Adeem G, Rubab A, Iqbal R. MAC protocols 802.11: A comparative study of throughput analysis and improved LEACH. In2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) 2020 Jun 24 (pp. 421-426). IEEE.

[5] Pamarthi S, Narmadha R. Adaptive Key Management-Based Cryptographic Algorithm for Privacy Preservation in Wireless Mobile

Adhoc Networks for IoT Applications. Wireless Personal Communications. 2022 May;124(1):349-76.

[6] Ebazadeh Y, Fotohi R. A reliable and secure method for network-layer attack discovery and elimination in mobile ad-hoc networks based on a probabilistic threshold. Security and Privacy. 2022 Jan;5(1):e183.

[7] Ganesh SS, Ravi G. A stable link connectivity-based data communication through neighbour node using traffic-less path in MANET. International Journal of Vehicle Information and Communication Systems. 2020;5(1):72-89.

[8] M. G. El-Hadidi and M. A. Azer, "Traffic Analysis for Real Time Applications and its Effect on QoS in MANETs," 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), 2021, pp. 155-160, doi: 10.1109/MIUCC52538.2021.9447611.

[9] M. A. Azer, N. G. Saad, “Prevention of Multiple Coordinated Jellyfish Attacks in Mobile Ad Hoc Networks” International Journal of Computer Applications. 2015 Jan 1;120(20).

[10] Hemalatha S, Kshirsagar PR, Manoharan H, Vasantha Gowri N, Vani A, Qaiyum S, Vijayakumar P, Tirth V, Haleem SL, Chakrabarti P, Teressa DM. Novel Link Establishment Communication Scheme against Selfish Attack Using Node Reward with Trust Level Evaluation Algorithm in MANET. Wireless Communications and Mobile Computing. 2022 May 6;2022.

[11] Bondada P, Samanta D, Kaur M, Lee HN. Data Security-Based Routing in MANETs Using Key Management Mechanism. Applied Sciences. 2022 Jan;12(3):1041.

[12] Almalki FA, Soufiene BO. EPPDA: an efficient and privacy-preserving data aggregation scheme with authentication and authorization for IoT-based healthcare applications. Wireless Communications and Mobile Computing. 2021 Mar 9;2021.

[13] Sultana T, Mohammad AA, Gupta N. Importance of the Considering Bottleneck Intermediate Node During the Intrusion Detection in MANET. In Research in Intelligent and Computing in Engineering 2021 (pp. 205-213). Springer.

[14] Shaukat K, Alam TM, Luo S, Shabbir S, Hameed IA, Li J, Abbas SK, Javed U. A review of time-series anomaly detection techniques: A step to future perspectives. In Future of Information and Communication Conference 2021 Apr 29 (pp. 865-877). Springer, Cham.

[15] Kumar S, Dutta K. Intrusion detection in mobile ad hoc networks: techniques, systems, and future challenges. Security and Communication Networks. 2016 Sep 25;9(14):2484-556.

[16] Singh S, Sharma S, Sharma S, Alfarraj O, Yoon B, Tolba A. Intrusion Detection System based Security Mechanism for Vehicular ad-hoc Networks for Industrial IoT. IEEE Consumer Electronics Magazine. 2021 Dec 28.

[17] Khan K, Mehmood A, Khan S, Khan MA, Iqbal Z, Mashwani WK. A survey on intrusion detection and prevention in wireless ad-hoc networks. Journal of Systems Architecture. 2020 May 1;105:101701.

[18] Li W, Meng W, Kwok LF. Surveying Trust-based Collaborative Intrusion Detection: State-of-the-Art, Challenges and Future Directions. IEEE Communications Surveys & Tutorials. 2021 Dec 28.

[19] Ramesh S, Yaashuwanth C, Prathibanandhi K, Basha AR, Jayasankar T. An optimized deep neural network based DoS attack detection in wireless video sensor network. Journal of Ambient Intelligence and Humanized Computing. 2021 Jan 2:1-4.

[20] Marathe NR, Shinde SK. Improved itca method to mitigate network-layer attack in manet. InData Communication and Networks 2020 (pp. 245-253). Springer, Singapore.

[21] Sinha S, Paul A. Neuro-fuzzy based intrusion detection system for wireless sensor network. Wireless Personal Communications. 2020 Sep;114(1):835-51.

[22] S.R. Snapp, J. Brentano, G.V. Dias, T.L. Goan, L.T. Heberlein, C.L. Ho, K.N. Levitt, "DIDS (distributed intrusion detection system)-motivation, architecture, and an early prototype". In Proceedings of the 14th national computer security conference 1991 (Vol. 1, pp. 167-176).

[23] F. Cuppens, and A. Miege, "Alert correlation in a cooperative intrusion detection framework". In Security and privacy, 2002. Proceedings. 2002 IEEE symposium on (pp. 202-215). IEEE.

[24] P. Kannadiga, and M. Zulkernine, "DIDMA: A distributed intrusion detection system using mobile agents". In Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2005 (pp. 238-245). IEEE.

[25] Staniford-Chen S, Cheung S, Crawford R, Dilger M, Frank J, Hoagland J, Levitt K, Wee C, Yip R, Zerkle D. GrIDS-a graph based intrusion detection system for large networks. InProceedings of the 19th national information systems security conference 1996 Oct 22 (Vol. 1, pp. 361-370).

[26] Zhang Z, Li J, Manikopoulos CN, Jorgenson J, Ucles J. HIDE: a hierarchical network intrusion detection system using statistical preprocessing and neural network classification. InProc. IEEE Workshop on Information Assurance and Security 2001 Jun 5 (pp. 85-90).

[27] Axelsson S. Intrusion detection systems: A survey and taxonomy. Technical report; 2000 Mar 14.

[28] Janakiraman R, Waldvogel M, Zhang Q. Indra: A peer-to-peer approach to network intrusion detection and prevention. In Enabling Technologies: Infrastructure for Collaborative Enterprises, 2003. WET ICE 2003. Proceedings. Twelfth IEEE International Workshops on 2003 Jun 9 (pp. 226-231). IEEE.

[29] Zhou, C.V., Karunasekera, S. and Leckie, C., 2007, May. Evaluation of a decentralized architecture for large-scale collaborative intrusion detection. In Integrated Network Management, 2007. IM'07. 10th IFIP/IEEE International Symposium on (pp. 80-89). IEEE.

[30] M. Raya, J.P. Hubaux, and I. Aad, " DOMINO: a system to detect greedy behavior in IEEE 802.11 hotspots. In Proceedings of the 2nd international conference on Mobile systems, applications, and services 2004, (pp. 84-97). ACM.

[31] M. Roesch, "Snort: Lightweight intrusion detection for networks. In Lisa 1999 (Vol. 99, No. 1, pp. 229-238).

[32] Fung CJ. Collaborative Intrusion Detection Networks and Insider Attacks. JoWUA. 2011 Mar;2(1):63-74.

[33] Kruegel, C., Toth, T., Kerer, C.: Decentralized Event Correlation for Intrusion Detection. In: International Conference on Information Security and Cryptology (2002).

[34] Maciá-Pérez F, Mora-Gimeno FJ, Marcos-Jorquera D, Gil-Martínez-Abarca JA, Ramos-Morillo H, Lorenzo-Fonseca I. Network intrusion detection system embedded on a smart sensor. IEEE Transactions on Industrial Electronics. 2011 Mar;58(3):722-32.

[35] Albers P, Camp O, Percher JM, Jouga B, Me L, Puttini RS. Security in Ad Hoc Networks: a General Intrusion Detection Architecture Enhancing Trust Based Approaches. In Wireless Information Systems 2002 Apr 3 (pp. 1-12).

[36] Gu G, Fogla P, Dagon D, Lee W, Skorić B. Measuring intrusion detection capability: an information-theoretic approach. In Proceedings of the 2006 ACM Symposium on Information, computer and communications security 2006 Mar 21 (pp. 90-101). ACM.

[37] Kumar G. Evaluation metrics for intrusion detection systems-a study. International Journal of Computer Science and Mobile Applications, II. 2014 Nov.

[38] F. Alam TM, Shaukat K, Hameed IA, Khan WA, Sarwar MU, Iqbal F, Luo S. A novel framework for prognostic factors identification of malignant mesothelioma through association rule mining. Biomedical Signal Processing and Control. 2021 Jul 1;68:102726.

[39] MIT Lincoln Laboratory: DARPA Intrusion Detection Evaluation, https://www.ll.mit.edu/ideval/data/2000data.html.

[40] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed and M. Xu, "A Survey on Machine Learning Techniques for Cyber Security in the Last Decade," in IEEE Access, vol. 8, pp. 222310-222354, 2020, doi: 10.1109/ACCESS.2020.3041951.

[41] Shaukat K, Luo S, Varadharajan V, Hameed IA, Chen S, Liu D, Li J. Performance comparison and current challenges of using machine learning techniques in cybersecurity. Energies. 2020 Jan;13(10):2509.

[42] Shaukat K, Luo S, Chen S, Liu D. Cyber threat detection using machine learning techniques: A performance evaluation perspective. In2020 International Conference on Cyber Warfare and Security (ICCWS) 2020 Oct 20 (pp. 1-6). IEEE.

Data Collection Method for Energy Storage Device of Distributed Integrated Energy Station based on Double Decision Tree

Hao Chen*, Guilian Wu, Linyao Zhang, Jieyun Zheng
Economic and Technology Institute, State Grid Fujian Electric Power Co.,Ltd., Fuzhou Fujian 350013, China

***Abstract*—The distributed integrated energy station includes an electric energy storage device, heat storage device, cold storage device and other devices. Aiming at the problem of low data acquisition accuracy of energy storage device caused by using a single sensor or acquisition scheme in the existing methods, a new data acquisition method of energy storage device of distributed integrated energy station is designed based on double decision tree algorithm. The data acquisition process of double decision tree algorithm is constructed. On the basis of the process, the mathematical models of electric energy storage device, heat storage device, cold storage device and hybrid energy storage device are established. Then the double decision tree algorithm is used to solve the constructed model, and the acquisition pseudo code is given. So far, the data acquisition of distributed integrated energy station energy storage device based on double decision tree has been completed. The results of case analysis show that the accuracy of this method is higher than 98%, and the collection time is less than 30 ms.**

***Keywords*—Double decision tree; distributed; integrated; energy station; energy storage device; data collection**

I. Introduction

The energy storage devices of integrated energy stations such as photovoltaic, electric energy storage, cold and heat energy storage, and natural gas combined cooling, heating and power supply can improve the reliability and energy utilization efficiency of the power supply system [1-2], and reduce system line losses. However, each energy system usually operates independently, resulting in the inability of many energy systems to coordinate effectively, which is prone to safety hazards and low energy utilization problems [3-4]. Therefore, the efficient collection of energy storage device data in distributed integrated energy stations is of great significance to its rapid development [5].

A decision tree algorithm is a method of constructing a decision tree for preliminary screening data based on data attribute information. The algorithm has higher learning performance, higher computing precision and higher computing efficiency. This algorithm has been widely used in many fields [6]. However, the decision tree operation is more complex and the parallel computing capability is low. Based on this, the dual decision tree algorithm came into being. It has the advantage of enhancing the parallel computing capability of the decision tree, and can be applied to collect the data from the energy storage device of the distributed energy station. Among them, the C4.5 algorithm is an efficient algorithm in the decision tree algorithm. Using the C4.5 algorithm to generate a decision tree helps to accurately select data collection points from the data source and facilitate subsequent data processing [7]. The dual decision tree algorithm can use decision trees to filter the data required by the user from the data set. After it establishes the data set, it is processed again through another decision tree to achieve high-quality and high-efficiency data collection.

Li Junnan et al. studied the collection and application of power energy big data based on the big data cloud platform [8]. This method utilizes the big data cloud platform to realize the collection of power and energy big data. Du Peng et al. studied a wide-area data acquisition scheme based on the power dispatching data network [9]. Certain results have been obtained using the wide area method of data collection in the power dispatching data network. However, the above two methods are applied to the data collection of energy storage devices in distributed integrated energy stations. Because the device needs to collect data on four aspects of electricity storage, cold storage, heat storage and hybrid energy storage during the data collection process. As a result, the acquisition accuracy is low and the time is long.

In order to solve the above problems, this paper studies the data acquisition method of the energy storage device of the distributed integrated energy station based on the double decision tree. It establishes the mathematical model of each energy storage device in the distributed integrated energy station. The dual decision tree algorithm is applied to the data collection of the energy storage device of the distributed comprehensive energy station, which improves the data collection effect of the energy storage device of the distributed comprehensive energy station.

II. Data Collection of Energy Storage Devices in Distributed Integrated Energy Stations

In this paper, in the process of researching the data acquisition method of the energy storage device of the distributed integrated energy station, the data sampling idea based on the double decision tree is first designed. Then, on this basis, the mathematical model of each energy storage device in the distributed integrated energy station is established. Finally, the dual decision tree algorithm is applied to the data collection of the energy storage device of the distributed integrated energy station, and the pseudo code is generated.

*Corresponding Author.

A. Data Collection Idea of Double Decision Tree Algorithm

The process of the data collection method of the energy storage device of the distributed integrated energy station based on the double decision tree is as follows:

In the sample, a decision tree for initially collecting the original data of the energy storage device is established through the C4.5 algorithm. It uses the established decision tree to filter and sort out the original data, realize the preliminary screening of the data, and make the re-collected data samples show a balanced distribution state [10]. The established decision tree can be reused, reducing data training and collection time. A second tree is established based on the data contained in the first established tree, and the collected data is further analyzed. The specific process is shown in Fig. 1.

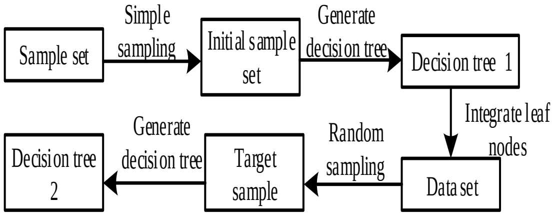


Fig. 1. Double Decision Tree Collection Process.

In Fig. 1, the sample set is the energy storage device data of the target distributed integrated energy station. It includes historical operation data of the energy storage device and corresponding file information.

On the basis of the acquisition process of the double decision tree in Fig. 1, a mathematical model of the energy storage device of the distributed integrated energy station is established.

B. Constructing Mathematical Model of Energy Storage Device in Distributed Integrated Energy Station

Distributed integrated energy stations usually include three loads of heat, cooling and electricity. Therefore, before collecting the data of the energy storage device of the distributed integrated energy station, the mathematical model of the energy storage device of the distributed integrated energy station is established from the aspects of electricity storage, heat storage, cold storage and hybrid energy storage.

1) Electric energy storage device model: The establishment of the model of the electric energy storage device in the distributed integrated energy station should focus on considering the electric energy capacity constraints, the maximum power, the complementary constraints and the charging and discharging efficiency [11]. In the process of model establishment, the energy consumption of the electric energy storage device caused by the change in time during the operation of the electric energy storage device is not considered.

When the time is set as t, the charging power and discharging power of the electric energy storage device are represented by $P_{char}(t)$ and $P_{dis}(t)$ respectively, and $W_E(t)$ is the remaining power of the energy storage device. In this paper, the mathematical model of the electric energy storage device of the distributed integrated energy station is established as follows:

$$\begin{cases} 0 \le P_{dis}(t) \le P_{\max}, 0 \le P_{char}(t) \le P_{\max} \\ W_E(t) = W_E(0) + \int_0^t \left[\eta_c P_{char}(t) - P_{dis}(t) / \eta_d \right] dt \\ W_{E_{\min}} \le W_E(t) \le W_{E_{\max}} \\ P_{dis}(t) P_{char}(t) = 0 \end{cases} \tag{1}$$

In formula (1), η_c represents the charging efficiency of the electric energy storage device. η_d represents the discharge efficiency of the electric energy storage device. $W_E(0)$ and $P_{\max}$ represent the initial remaining power and the maximum charging and discharging power of the electric energy storage device, respectively. $W_{E_{\min}}$, $W_{E_{\max}}$ and $P_{dis}(t)P_{char}(t) = 0$ respectively represent the remaining power operation area of the electric energy storage device and the complementary constraint, and the complementary constraint is used to limit the state unity of the electric energy storage device.

2) Case analysis Model of heat and cold storage device: The distributed integrated energy station adopts cold storage and thermal storage to realize sensible heat energy storage and phase change energy storage. Sensible heat energy storage is a low-cost energy storage method. Phase change energy storage realizes energy storage by storing and releasing heat in the process of changing the physical state of materials. It has a higher cost [12], but the temperature fluctuation is small. The phase change energy storage method is selected as the energy storage method of the heat storage and cold storage devices of the distributed integrated energy station. This method has a high energy storage density, and when the heat storage and temperature difference are the same, it can ensure that the heat storage of the phase change material is only one-quarter to one-fifth of the sensible heat material. The temperature during energy storage and release can be controlled and kept constant.

Both cold storage and heat storage devices have energy consumption characteristics when storing and releasing energy. And the energy has a certain dissipation with time.

The model formula of the heat storage (cold storage) device of the distributed integrated energy station is as follows:

$$\begin{cases} 0 \le P_{TI}(t) \le P_{\max}, 0 \le P_{TO}(t) \le P_{\max} \\ W_P(t) = \eta_T W_P(t-1) + \eta_{TI} P_{TI}(t) - P_{TO}(t) / \eta_{TO} \\ 0 \le W_P(t) \le R_{HS} \\ P_{TI}(t) P_{TO}(t) = 0 \end{cases} \tag{2}$$

In formula (2), $W_P(t)$ and $1-\eta_T$ respectively represent the residual heat of the heat storage (cold storage) device at time t and the loss rate of the heat storage (cold storage) device per unit time. η_{TI} represents the heat storage (cold storage) efficiency of the heat storage (cold storage) device. η_{TO} represents the heat (cold) efficiency of the heat storage (cold storage) device.

R_{HS} and $P_{\max}$ represent the maximum capacity and maximum heat release (cold) power of the heat storage (cold storage) device, respectively, $P_{TI}(t)$ represents the heat storage (cold storage) power of the heat storage (cold storage) device when the time t is time; $P_{TO}(t)$ represents the heat storage (cold storage) power when the time t is Cold storage) device to release heat (cold) power.

The calculation formulas of the two heat storage and cooling efficiencies are as follows:

$$\eta_I = \frac{1.16G(\mathrm{T}\,2-\mathrm{T}1)\times 100\%}{Pt_1/60} \tag{3}$$

$$\eta_O = \frac{1.16G(\mathrm{T}1-\mathrm{T}\,2)\times 100\%}{Pt_2/60} \tag{4}$$

In the formula, the total amount of hot water collected after 90s of water supply to the end of heat storage or cold storage is described by G. The power is denoted by P, and the collected hot water and cold water time are denoted by t_1 and t_2, respectively. The mean value of outlet water temperature and the value of cold water temperature are described by $\mathrm{T}1$ and $\mathrm{T}\,2$, respectively.

1) Hybrid energy storage system model: The internal energy of the energy storage system of the distributed integrated energy station is in a complementary state [13]. Compressed air energy storage and molten salt heat storage are used in molten salt heat storage non-supplementary combustion compressed air energy storage systems. Combining multiple energy stores can improve the economics of energy storage systems.

The turbine inlet is heated by the heat of the molten salt heat storage system of the hybrid energy storage device. Convert thermal energy into electrical energy, and set an electrothermal device to convert electrical energy into thermal energy [14]. It realizes the four-quadrant operation of the hybrid energy storage device of the distributed integrated energy station for two types of energy flows, electric energy and thermal energy.

According to the four-quadrant operation principle of the hybrid energy storage device, P_{in} is represented by P_o and the electric energy input power and output power of the hybrid energy storage device, respectively. H_{in} and H_o represent the electric energy input power and output power of the hybrid energy storage device, respectively. τ_1 represents the energy input time interval. τ_2 represents the energy output time interval, and the model of the hybrid energy storage device can be obtained as follows:

$$\begin{bmatrix} P_o(t) \\ H_o(t) \end{bmatrix} = \begin{bmatrix} \eta_{Com}\eta_{Tur} & \eta_{Tur}\eta_{Hs}\eta_{H-E} \\ \eta_{Heat}(1-\eta_T\tau_1) & \eta_{Hs}(1-\eta_T\tau_2) \end{bmatrix} \begin{bmatrix} P_{in}(t-\tau_1) \\ H_{in}(t-\tau_2) \end{bmatrix} \tag{5}$$

In the formula, η_{Com} represents the compressor efficiency. η_{Tur} stands for turbine efficiency. η_{Hs} and η_{H-E} represent heat storage efficiency and turbine inlet air heating efficiency, respectively. η_{Heat} represents the heater efficiency.

C. Data Acquisition for Energy Storage Equipment of Distributed Comprehensive Energy Station

The dual decision tree algorithm is applied to the established model of each energy storage device of the distributed integrated energy station. The data collection of the energy storage device is realized by using the double decision tree algorithm.

The decision tree algorithm is a typical inductive reasoning algorithm. The decision tree algorithm needs to clarify the attributes of each node of the tree and obtain the required attribute data. In order to improve the operation performance of the decision tree algorithm, the concept of information gain is introduced into the decision tree algorithm. We use the amount of information gained in the operation process to clarify the test attributes required for each node of the decision tree [15], and the decision tree algorithm is usually used to select attributes with a larger number in the operation process.

The C4.5 algorithm is a decision tree method that replaces the information gain of the attribute classification level evaluation index with the information gain rate. This method can effectively improve the defect that the traditional decision tree algorithm is limited to local optimization. The C4.5 algorithm uses the automatic discretization method to process the attributes with continuous values, and avoids the over-learning of the decision tree by pruning the decision tree. Algorithm C4.5 builds a decision tree using information that is relevant to the collection and classification. Let L represent the case set, C_i represent the case sample class label, and $i=1,2,\cdots,n$, and the information entropy formula for data collection can be obtained as follows:

$$I(L) = -\sum_{i=1}^{n} \frac{F(C_i,L)}{|L|}\log_2\frac{F(C_i,L)}{|L|} \tag{6}$$

In the formula, $|L|$ and $F(C_i,L)$ represent the number of samples in the case set L and the number of cases belonging to the C category in the case set A, respectively.

In the case where the number of values k exists in the selected attribute X . According to the probability of each information obtained from the training set, the formula for the conditional entropy of the decision tree is formed as follows:

$$E_x = \sum_{i=1}^{k} \frac{|L_i|}{|L|} I(L_i) \tag{7}$$

In the formula, $|L_i|$ is the number of cases of various subtrees in the attribute X , and the formula for obtaining mutual information is as follows:

$$G(X) = I(L) - E_x \tag{8}$$

Algorithm C4.5 selects the heuristic search extended attribute score, and the extended attribute selects the attribute with the largest information gain. The heuristic method can be effectively applied to the normalization process and in the presence of different attribute values, the attributes that reflect high-quality information gain can be selected. It obtains the attribute information gain rate formula as follows:

$$g_\tau(X) = \frac{G(X)}{Z(X)} \tag{9}$$

Branch the energy storage device data by the value of attribute X to obtain the $Z(X)$ value of the dataset, $Z(X) = -\sum_{i=1}^{k} \frac{|L_i|}{|L|} \log_2 \left(\frac{|L_i|}{|L|} \right)$. After completing the above calculations, based on the flow of the dual decision tree data collection method in Fig. 1, simple random sampling is performed on the original data set to obtain an initialization sample B . It applies the C4.5 algorithm to build the first decision tree within the initialization sample. After traversing all the leaf nodes of the first decision tree, the corresponding samples are stored in the sample data set, and the data is sent to the next data set $B_i\{i = 1, 2, \mathrm{L}\ \ , m\}$ until the set requirements are met. Sampling stops after all data traversal is completed. Randomly draw samples from the data set obtained by the first decision tree, and use $Z * \frac{|B_i|}{\sum_i^m B_i}$ to represent the number of samples from the data set i , where Z and $|B_i|$ represent the number of target samples and the number of samples in the dataset i , respectively; The target sample Z_t is all the samples extracted, and the second decision tree is generated by using the target sample Z_t that has been extracted. Through the second decision tree, the final collection of the data of the energy storage device of the distributed integrated energy station is realized.

The pseudo code for generating the data collection method of the energy storage device of the distributed integrated energy station with the dual decision tree is shown in Fig. 2.

```
Input: dataset Z
Output: selected data
Initialize Tree←NULL;
If Z is empty or meet other end conditions, then terminate;
End if
For all attributes in the dataset $a$ do
    Calculate the information gain rate;
End For
$a_{best}$← the attribute with the largest information gain rate;
Tree←$a_{best}$ is the root decision node;
$Zv$←Sub-dataset based on $a_{best}$ division;
For all $Zv$ do
    $Treev$←C4.5($Zv$)
    Add $Treev$ to the corresponding branch of $Tree$;
End for
Return $Tree$
Find Last Child Node(Tree Node $node$ , ArrayList $Zi$($i$=1,2,…,$m$));
If no child node then
    do $Zi$.Add($node$);
    If the number of $Zi$ samples < initial limit value then
        do $i$ ++;
    End if
Else
    For Tree Node $n$ in $node$. Child Nodes do
        this  Find Last Child Node ($n$,$Zi$)
    End for
For all $Zi$ do
    Random sampling, the number of samples is $Z * \frac{|B_i|}{\sum_i^m B_i}$ ;
    Integrate into the target sample set $Zt$
End for
Input dataset $Zt$;
Initialize $Tree$←NULL
Terminate if $Zt$ is empty or encounter other end conditions
End if
For all attributes in the dataset $a$ do
    Calculate the information gain rate;
End for
$a_{best}$← The attribute with the largest information gain rate;
$Tree$←$a_{best}$ is the root decision node;
$Zv$←Sub-dataset based on $a_{best}$ division;
For all $Zv$ do
    Add the $Tree$ to the corresponding branch of the $Tree$
    $Treev$ ← C4.5 ($Zv$)
End for
Return $Tree$
```

Fig. 2. Data Acquisition Pseudocode.

So far, the design of the data acquisition method for the energy storage device of the distributed integrated energy station based on the double decision tree is completed. The next step will verify the effectiveness of the proposed method through case analysis.

III. Case Analysis

In order to validate the data acquisition method of distributed comprehensive energy storage device based on double decision tree is effective. A distributed integrated energy station in a certain industrial park is selected as the experimental object. The distributed comprehensive energy station includes three electric energy storage devices, two heat storage devices, two cold storage devices, and one compressed air hybrid energy storage device. The method of this paper is used to collect the data of each energy storage device of the distributed integrated energy station, and the validity of the data collection method of this paper is verified.

A. Energy Storage Device Data Acquisition Test

We set the sampling time to 20s, and the actual sampling interval is 2s. The LMG671 conventional broadband power detection instrument and the designed data acquisition method of the energy storage device of the distributed integrated energy station based on the double decision tree are used to collect the power of each energy storage device. The results within the statistical sampling time are shown in Tables I to III.

TABLE I. ACTUAL POWER（KW）

Sampling time/s	Electric energy storage device 1	Electric energy storage device 2	Electric energy storage device 3	Heat storage device 1	Heat storage device 2	Cold storage device1	Cold storage device2	Hybrid energy storage device
2	1626	2217	2366	3127	2856	2535	2685	2353
4	1653	2234	2385	3153	2846	2517	2676	2342
6	1624	2285	2350	3148	2862	2538	2648	2364
8	1626	2265	2358	3122	2816	2583	2649	2313
10	1653	2235	2363	3117	2836	2568	2636	2379
12	1635	2220	2349	3155	2866	2531	2657	2365
14	1688	2296	2361	3194	2839	2519	2687	2393
16	1626	2248	2348	3159	2863	2564	2675	2354
18	1687	2236	2317	3176	2818	2576	2632	2359
20	1647	2265	2376	3150	2837	2599	2686	2341

TABLE II. LMG671 CONVENTIONAL POWER ACQUISITION RESULTS（KW）

Sampling time/s	Electric energy storage device 1	Electric energy storage device 2	Electric energy storage device 3	Heat storage device 1	Heat storage device 2	Cold storage device1	Cold storage device2	Hybrid energy storage device
2	1628	2220	2369	3131	2859	2539	2682	2356
4	1650	2237	2388	3150	2842	2520	2672	2346
6	1627	2281	2349	3144	2866	2541	2651	2367
8	1623	2269	2354	3126	2813	2586	2653	2317
10	1657	2231	2367	3113	2840	2564	2632	2382
12	1637	2223	2346	3159	2862	2534	2653	2362
14	1682	2293	2365	3191	2836	2519	2682	2397
16	1623	2243	2345	3163	2860	2561	2672	2350
18	1681	2239	2321	3172	2822	2571	2628	2363
20	1642	2269	2372	3154	2831	2602	2682	2338

TABLE III. THE POWER ACQUISITION RESULTS OF THE METHOD IN THIS PAPER（KW）

Sampling time/s	Electric energy storage device 1	Electric energy storage device 2	Electric energy storage device 3	Heat storage device 1	Heat storage device 2	Cold storage device 1	Cold storage device 2	Hybrid energy storage device
2	1625	2215	2364	3125	2854	2534	2685	2352
4	1652	2234	2385	3152	2843	2516	2675	2341
6	1623	2285	2349	3147	2861	2537	2647	2364
8	1624	2264	2357	3124	2814	2584	2651	2315
10	1652	2234	2361	3117	2834	2567	2635	2378
12	1634	2218	2347	3152	2864	2534	2654	2364
14	1685	2296	2359	3194	2837	2517	2685	2391
16	1625	2247	2347	3158	2861	2564	2675	2354
18	1685	2234	2315	3175	2816	2574	2634	2361
20	1647	2264	2374	3149	2837	2598	2684	2341

Comparing and analyzing the three tables above, it can be seen that the power value of the energy storage device collected by the method in this paper is basically the same as the actual value. However, the difference between the power value detected by the conventional detection device and the actual value is higher than the difference between the method in this paper and the actual value. It shows that compared with the conventional detection device, the power value of the energy storage device collected by the method in this paper is more accurate.

On the basis of the above collection results, in order to further verify the data collection performance of the method in this paper, accuracy is selected as the evaluation index to evaluate the data collection performance. The accuracy of the power data collected by the method of this paper, the big data cloud platform method [8] and the wide-area method [9] is collected for the 8 energy storage devices of the distributed integrated energy station. In order to visually demonstrate the data acquisition performance of the method in this paper, the MATLAB tool is used to generate the comparison results of the power acquisition accuracy of the energy storage device shown in Fig. 3.

From the experimental results in Fig. 3, it can be seen that the power collection accuracy of the distributed integrated energy station energy storage device power using the method in this paper is higher than that of the big data cloud platform method and the wide area method. The accuracy of the method in this paper to collect the data of the energy storage device of the distributed integrated energy station is higher than 98%. The experimental results effectively verify that the method in this paper has high data acquisition performance. The main reason is that the method in this paper uses the double decision tree algorithm to effectively improve the data collection performance through two decision tree processing. It has high data acquisition accuracy and high practicability of data acquisition of energy storage equipment in distributed integrated energy stations.

B. *Real-time Test of Data Acquisition of Energy Storage Device*

The real-time data collection of each energy storage device in a distributed integrated energy station affects the normal operation of the energy storage device. Set the attribute data to be collected to 2500. The method of this paper is used to collect the average collection time of each sample of the power, DC current and DC voltage of the energy storage equipment of the distributed integrated energy station. The statistical results are shown in Tables IV to VI.

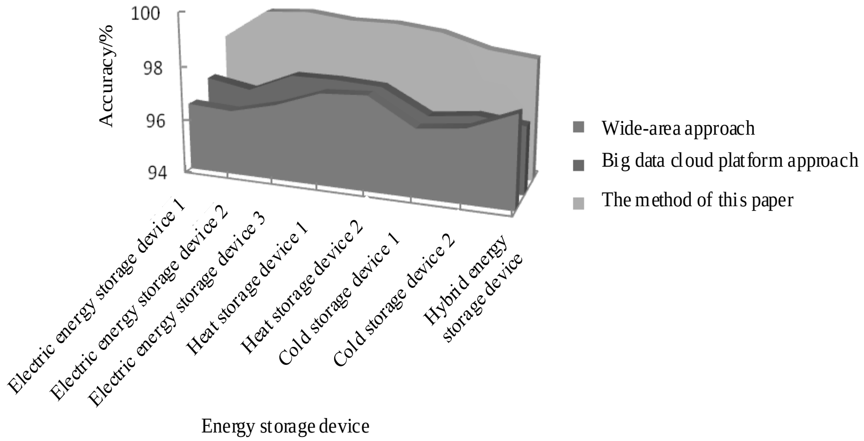


Fig. 3. Accuracy of Power Collection of Energy Storage Devices.

TABLE IV. ENERGY STORAGE DEVICE POWER COLLECTION RESULTS

Attribute data/pcs	Electric energy storage device 1/kW	Electric energy storage device 2/kW	Electric energy storage device 3/kW	Heat storage device1/kW	Heat storage device2/kW	Cold storage device1/kW	Cold storage device2/kW	Hybrid energy storage device/kW
500	1625	2215	2364	3125	2854	2534	2685	2352
1000	1652	2234	2385	3152	2843	2516	2675	2341
1500	1623	2285	2349	3147	2861	2537	2647	2364
2000	1624	2264	2357	3124	2814	2584	2651	2315
2500	1652	2234	2361	3117	2834	2567	2635	2378

TABLE V. DC VOLTAGE ACQUISITION RESULTS OF ENERGY STORAGE DEVICE

Sampling time/s	Electric energy storage device 1/V	Electric energy storage device 2/V	Electric energy storage device 3/V	Heat storage device1/V	Heat storage device2/V	Cold storage device1/V	Cold storage device2/V	Hybrid energy storage device/V
500	125.4	135.6	151.2	141.5	131.2	128.5	171.5	161.9
1000	124.2	134.5	150.4	141.6	131.5	128.6	170.8	160.24
1500	125.2	135.2	151.2	142.2	131.6	128.9	171.2	161.3
2000	125.4	134.4	151.3	141.6	131.8	128.7	171.4	161.8
2500	125.5	135.5	151.4	142.5	131.7	128.3	171.2	161.5

TABLE VI. DC CURRENT ACQUISITION RESULTS OF ENERGY STORAGE DEVICE

Sampling time/s	Electric energy storage device 1/A	Electric energy storage device 2/A	Electric energy storage device 3/A	Heat storage device1/A	Heat storage device2/A	Cold storage device1/A	Cold storage device2/A	Hybrid energy storage device/A
500	5.75	6.35	6.25	5.64	6.35	9.52	7.52	8.25
1000	5.35	6.52	6.34	5.28	6.85	9.45	7.46	8.64
1500	5.46	6.45	6.15	5.46	6.28	9.25	7.18	8.15
2000	5.27	6.85	6.28	5.31	6.45	9.34	7.64	8.34
2500	5.81	6.34	6.54	5.28	6.75	9.52	7.58	8.62

From the analysis of Tables IV to VI, it can be seen that the method in this paper can complete the target collection attribute data volume. And it counts the time required to collect different data volumes during the collection process, and draws its statistical results as Fig. 4.

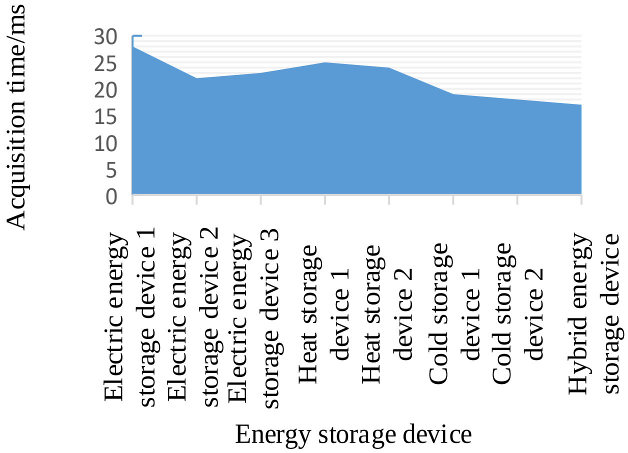


Fig. 4. Real-time Data Collection

The experimental results in Fig. 4 show that the power, DC current and DC voltage data collected by the method in this paper are all less than 30 ms. The reason for the short acquisition time of the method in this paper is that the decision tree for preliminary acquisition of the original data of the energy storage device is established by the C4.5 algorithm. It uses the established decision tree to filter and sort out the raw data. It realizes the preliminary screening of data and improves the collection efficiency. Using the method in this paper to collect various data from energy storage equipment in distributed integrated energy stations not only has high acquisition accuracy. This method has a high real-time acquisition, which again verifies the high data acquisition performance of the method in this paper. It can be applied to the practical application of data acquisition of energy storage devices in distributed integrated energy stations.

IV. CONCLUSION

Distributed comprehensive energy stations are an important direction for integrating and optimizing the energy Internet. In this paper, a mathematical model of each energy storage device in a distributed integrated energy station is established. It collects energy storage device data through a dual decision tree method and draws the following conclusions:

1) The accuracy of the proposed method is higher than 98%, and the collection time is less than 30ms, which can improve the data collection efficiency and collection accuracy. The method in this paper can obtain ideal data collection results in a short time, and the tree-like decision tree structure can intuitively reflect the status of the data to be collected and improve the accuracy of data collection. The method is applied to the data collection of energy storage devices in distributed integrated energy stations, and has high engineering practice value.

2) The proposed method improves the defect of inaccurate collection caused by too much useless information in massive data. It avoids unbalanced sampling space caused by differences in data distribution between different samples, and provides greater convenience for subsequent data analysis.

3) In this study, it is not considered that the electric energy storage device can provide heat and cooling data through the heat pump in the process of outputting electricity. The accuracy of power collection of energy storage devices needs to be further improved and studied.

REFERENCES

[1] Y. H. Jia and F. Zhang, “A bi-level optimal configuration of multiple storage in reginal integrated energy system with distribution wind power inclusion,” Renewable Energy Resources, 2019, 37(10), pp. 1524-1532.

[2] C. L. Wang, H. Liu and J. F. Gong, “Joint Scheduling of Different Energy Storage for Improving Wind Power Accommodation Ability in Integrated Community Energy System,” Electric Power Construction, 2018, 039(4), pp. 35-44.

[3] Z. H. Jiang, Y. Q. He, L. L. Cao, “Reconfiguration of distribution network with distributed generations and energy storing devices based on improved genetic algorithm,” Power System Protection and Control, 2018, 046(5), pp. 68-72.

[4] W. Xiong, Y. Q. Liu, W. H. Su, “Optimal configuration of multi-energy storage in regional integrated energy system considering multi-energy complementation,” Electric Power Automation Equipment, 2019, 39(1), pp. 124-132.

[5] Y. Lu, Y. Dai, W. Z. Ma, “Decentralized Dynamic Optimal Power Flow in Distribution Networks With Distributed Generation and Energy Storage Devices,” Power System Technology, 2019, 43(2), pp. 434-442.

[6] Y. W. Liu, Y. Hu, N. L. Tai, “Rule extraction method of operation and maintenance expert system for an intelligent substation based on the decision tree,” Journal of Electric Power Science and Technology, 2019, 34(1), pp. 125-130.

[7] W. Q. Sun, Z. Li, Y. M. Tan, “Method of Power System Energy Storage Configuration Based on Flexibility Promotion,” Journal of System Simulation, 2018, 30(1), pp. 235-241.

[8] J. N. Li, W. Li, H. J. Li, “Research on big data acquisition and application of power energy based on big data cloud platform,” Electrical Measurement & Instrumentation, 2019, 56(12), pp. 104-109.

[9] P. Du, L. Yan, B. C. Gao, “Wide-area Data Acquisition Scheme Based on Power Dispatching Data Network,” Automation of Electric Power Systems, 2019, 43(13), pp. 156-161.

[10] L. Chen, H. X. Fei, H. L. Ding, “A data sampling method based on double decision tree,” Computer Engineering and Science, 2019, 41(01), pp. 134-139.

[11] L. P. Zhang, X. Ye, J. Wang, “Research on method for data collection based on distribute power generation,” Renewable Energy Resources, 2017, 35(8), pp. 1203-1207.

[12] T. F. Ma, J. Y. Wu, L. L.Hao, “Energy Flow Modeling and Optimal Operation Analysis of Micro Energy Grid Based on Energy Hub,” Power System Technology, 2018, 42(1), pp. 179-186.

[13] K. K. Gu, W. S. Hu, K. Zhao, “Design and implementation of a handheld device for power data acquisition and analysis based on mobile network,” Power System Protection and Control, 2018, 506(8), pp. 115-121.

[14] H. B. Kang, Y. N. Qu, L. Zhao, “Research on microgrid control technology with distributed power supply and energy storage device,” Chinese Journal of Power Sources, 2017, 41(4), pp. 627-629.

[15] L. L. Chen, L. H. Mu, X. F. Xu, “Influences of energy storage operational strategy and characteristic on microgrid reliability,” Electric Power Automation Equipment, 2017, 37(7), pp. 70-76.

Reduced False Alarm for Forest Fires Detection and Monitoring using Fuzzy Logic Algorithm

Maria Susan Anggreainy, Bimo Kurniawan, Felix Indra Kurniadi
Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

***Abstract*—The purpose of this research was to detect forest fires in efforts to minimize the incidence of false alarms. The research method used is divided into several stages, namely planning, analysis, design, implementation, testing, and maintenance. Arduino acts as a data collector on the field, which will later be used to detect forest fires and false alarms. Fuzzy logic is used as the essence of the algorithm and will provide a higher level of accuracy for forest fire detection and false alarms. In testing the fuzzy program, the fuzzy output between Arduino and the fuzzy on the monitoring dashboard has a small difference of 0.99%. It can be concluded; the application can minimize the occurrence of false alarms to reduce the user's workload.**

***Keywords*—False alarm; fuzzy logic; forest fires detection; sensor; microcontroller**

I. Introduction

Forests are the lungs of the world because forests will absorb carbon dioxide (CO2) in the air and release oxygen (O2) which is very useful and creates a balance of ecosystems for living things. Therefore, forest fires are something that must be prevented and minimized their impact. The primary sources for ecological degradation currently are the Forest Fires [1]. Huge losses and serious threats to ecosystems are a common consequence of forest fires [2]. Indonesia is one of the countries with the largest forest area in the world. With forested land area throughout Indonesia's mainland is 94.1 million hectares or 50.1% of the total land area. However, although the forest area owned by Indonesia is quite high, the number of forest and land fires (Karhutla) in Indonesia is also quite high. Based on data quoted from the Directorate of Land and Forest Fire Control, Ministry of Environment and Forestry of the Republic of Indonesia, forest, and land fires in Indonesia in 2019 alone reached 1,649,258 hectares [3]. Forest fires are one of the big problems in Indonesia because forest fires are very dangerous impact on the environment and people due to smoke and carbon emissions from fires [4]. The most technology to detect fire Hotspots have been using satellite imagery and then processed to determine the number of hotspots and their locations. Some drawbacks in this technology that in bad weather or cloudy then the satellite system cannot penetrate [5][6]. In this research, it is proposed sensor system that uses multiple sensors related to fire parameters, especially fires on peatlands with special specifications fire case. Common fire parameters such as temperature, smoke, fog and carbon dioxide are applied in this system then measure the indicator using a special sensor [7]. This high number of forest and land fires is the source of several problems. Among them are disturbing human health as well as flora and fauna, hindering community activities such as flights and school activities, destroying the earth with CO2 content released during forest and land fires, and the clean water crisis caused by water that should be filtered by rocks in the ground, instead being mixed with post-fire soil [8]. The number of forests and lands that burn every year in Indonesia is one of the reasons why early detection of forest and land fires can help humans in handling forest and land fires. Therefore, the proposed solution to this problem is by utilizing the technology that is currently popular, namely Arduino which is one of the Internet of Things (IoT) [9]. Currently IoT Devices and sensors enable monitoring various environmental variables, such as temperature, humidity, etc. [10]. Arduino based platform IoT enabled fire detector and monitoring system is the solution to this problem [11] [12]. To implement this research, we will use GSM which is used to deliver WhatsApp to the user via the number given in the simulation program. Based on data from the Indonesia IoT Forum, there are 400 million sensor devices installed. However, use in the agricultural sector is only 4 percent. This low number can be interpreted by the wide space for IoT innovation in the agricultural sector. This monitoring dashboard is used to improve response to forest fires. So that the impact of forest and land fires can be minimized as much as possible. previous research there are several research on algorithms to calculate the probability of fire occurrence [13], forest fire models and warnings based on GIS grid platform [14] studying fire alarms and based on methods based on infrared technology [15]. However, the shortcomings of this study have not been able to reduce false alarms.

II. Methodology

The method used in this research is an experimental method. The system is built through the design stage hardware, software, implementation, and testing.

A. Hardware Design

Hardware design carried out based on the block diagram in Fig. 1. System. It consists of input, process, and output. Block diagram of a forest fire detector prototype using a power bank as a resource for the microcontroller and sensors. Starting from the input or input, the input block diagram is in the form of smoke sensors, fire sensors and temperature sensors which are used as input from the tool to start the fire detection process or not. For the microcontroller used is Arduino Uno Rev3 which functions as the main processing component and components that can connect input with output. In the output section there is no component to print the output, but it can be seen from the website and WhatsApp notifications using the mobile phone of forest guards.

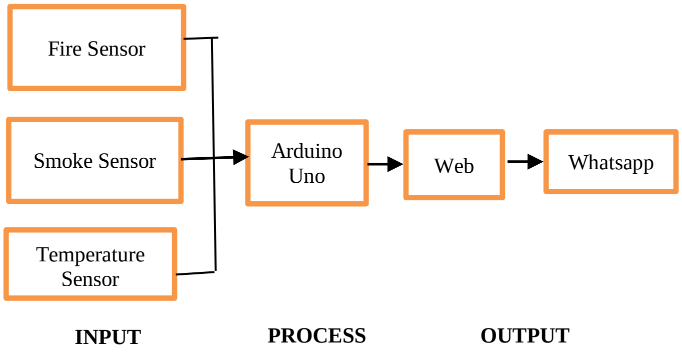


Fig. 1. Block Diagram.

There are three sensors used in this research, namely:

1) Smoke detection sensor: The smoke detection sensor used is the Gravity Analog Infrared CO2 Sensor. This sensor works by sucking air around the sensor, then the sensor will determine the gas concentration of the inhaled air [16]. The sensor will get the gas concentration result in Parts-per-million (ppm). For example, if the result is 1,000ppm CO2. So that means there are 1,000 CO molecules and 999,000 other gas molecules. Fig. 2 shows the physical form of the smoke detection sensor used.

Fig. 2. Gravity Analog Infrared CO2 Sensor.

2) Fire detection sensor: The fire detection sensor used is a 5 Channel Flame Sensor. This sensor uses infrared rays in 5 directions with a range of 120 degrees and in the wave range of 700 nm – 1100 nm, with a detection distance of less than 1 meter and a response time of about 15 microseconds. The fire detection sensor works by detecting infrared radiation generated by the fire, then the sensor converts the detection into analog and digital signals for further processing. Fig. 3 shows the physical form of the fire detection sensor used.

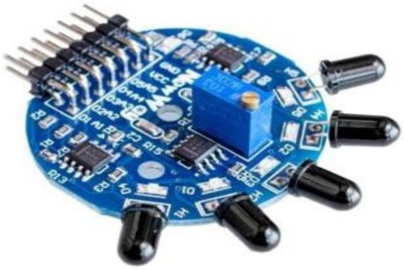


Fig. 3. Sensor Flame.

3) Temperature detection sensor: The temperature detection sensor used is the DS18B20 Temperature Sensor, this sensor has a sensor that is quite precise and does not require external components to work. This sensor can measure temperatures from -55°C to +125°C with an accuracy of ±0.5°C. Fig. 4 shows the physical form of the temperature detection sensor used.

Fig. 4. DS18B20 Temperature Sensor.

B. Software Design

The analysis is carried out in the form of observations of the surrounding environment. When a forest fire occurs it cannot be detected by forest rangers, forest guards only know that if a forest fire occurs, smoke will raise in the sky and the fire conditions are largely handled by the firefighters because the fire was too big. The detection process is carried out manually which makes fires difficult to control because the fires are getting bigger, burning forest areas, and the process of forest fires is relatively long. Therefore, an application is needed to monitor and detect forest fires that are useful for minimizing the incidence of false alarms and responding to forest fires handling. The business process of this application is that this application receives sensor data from the microcontroller, then the application will determine the fire status based on the sensor data received by utilizing fuzzy logic, and then the application will send a notification to the forest ranger via WhatsApp, if the fire status is a hazard, then this information also sent to firefighters. Application design business processes to monitor and detect forest fires can be seen in Fig. 5.

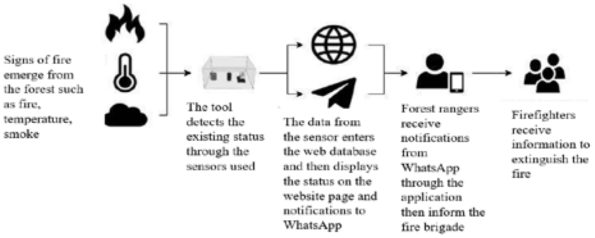

Signs of fire emerge from the forest such as fire, temperature, smoke
The tool detects the existing status through the sensors used
The data from the sensor enters the web database and then displays the status on the website page and notifications to WhatsApp
Forest rangers receive notifications from WhatsApp through the application then inform the fire brigade
Firefighters receive information to extinguish the fire

Fig. 5. Design Business Processes.

The flowchart of the system created is described by the sequence of processes in detail from one process to the next so that the system flows easy to understand. Fig. 6 shows the flowchart of the system that has been made the classification procedure begins with the receipt of sensor data sent by the microcontroller via the API. After that the application will process the data using fuzzy logic to determine the fire status. Fuzzy logic works by receiving data received from the sensor, then classify each data based on the membership function and generate the degree of membership of each sensor data. After that the value of the membership degree will be processed using the rule and will produce the fire status membership degree, which is then entered into the membership degree graph to determine the fire status. If the fire status is declared dangerous, then the application will send a notification message to all contacts WhatsApp stored in the database and displays the status of the fire as a hazard on the monitoring dashboard. Meanwhile, if the fire status is safe or alert, the application only displays the status of a safe or alert fire on the monitoring dashboard.

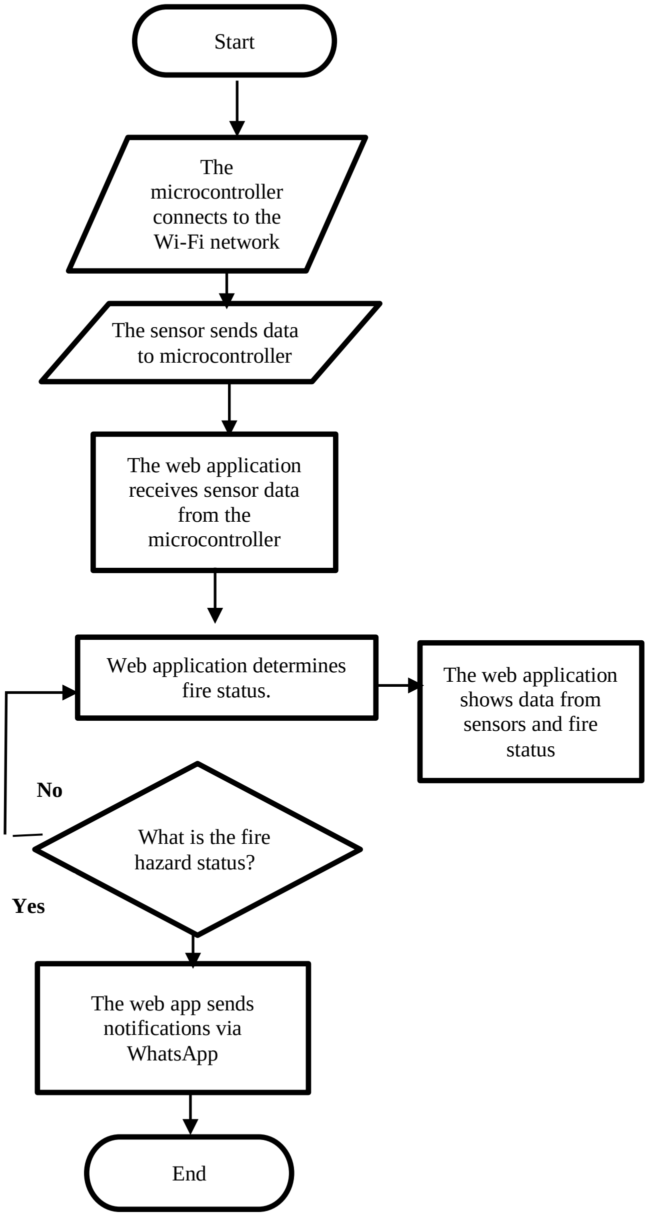

Start
The microcontroller connects to the Wi-Fi network
The sensor sends data to microcontroller
The web application receives sensor data from the microcontroller
Web application determines fire status.
The web application shows data from sensors and fire status
No
What is the fire hazard status?
Yes
The web app sends notifications via WhatsApp
End

Fig. 6. Flowchart System.

In Fig. 7, the fuzzification sub-system will process input data when doing data reading [17]. The data is in the form of firm values or crispy. In this research using Sugeno Fuzzy for inferencing [18]. The following are the stages of the process of Sugeno Fuzzy Inference:

1) Fuzzification: The fuzzification sub process will change the firm value that exists in the membership function or degree of membership. Based on the analysis carried out, the fire status is divided into three, namely, safe, caution, and danger. Where the three statuses are determined based on the variables of Temperature, Smoke, and Fire. Each of these variables has its own status classification. Temperatures are classified into three states, namely, Normal with temperatures less than or equal to 30°C, Warm with temperatures between 30°C and 60°C, and Hot with temperatures greater than or equal to 60°C.

The temperature membership set can be seen in Fig. 8, the membership set function is divided into three ranges, namely, normal [15, 30, 45], warm [30, 45, 60], and hot [45, 60, 75].

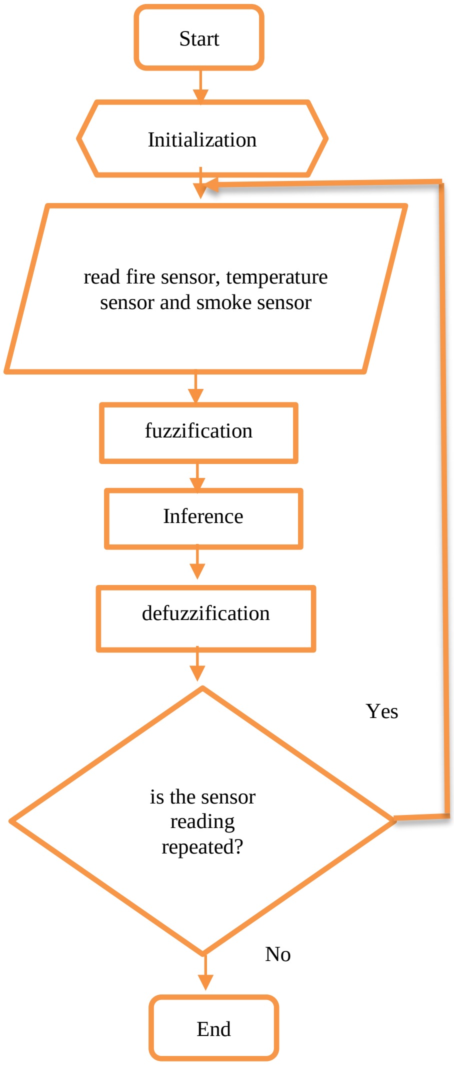

Start
Initialization
read fire sensor, temperature sensor and smoke sensor
fuzzification
Inference
defuzzification
Yes
is the sensor reading repeated?
No
End

Fig. 7. Fuzzy Design.

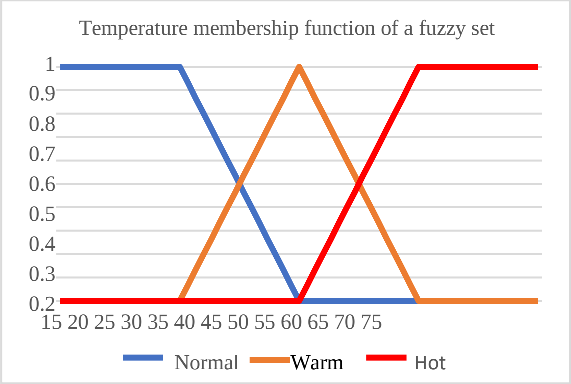


Fig. 8. Temperature Membership.

The smoke membership set can be seen in Fig. 9, in the membership set function is divided into three ranges, namely, thin [200, 600, 1000], medium [600, 1000, 1400], and dense [1000, 1400, 1800].

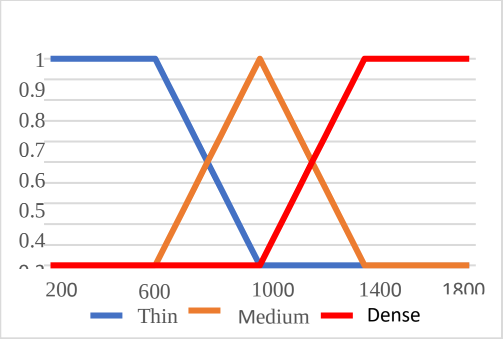


Fig. 9. Smoke Membership.

Fire membership set can be seen in Fig. 10, fire membership set is divided into two ranges, namely, true and false.

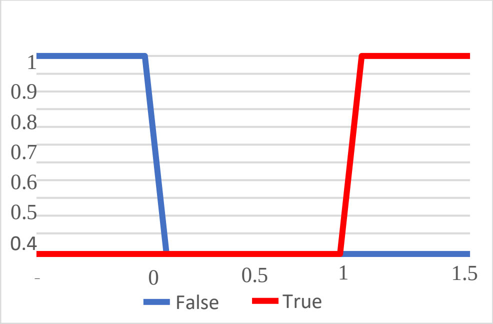


Fig. 10. Fire Membership.

The membership sets above will be used to get the output value by going through the implication stage.

2) Inference: At this stage each rule or rule in the fuzzy knowledge base will be associated with a fuzzy relation. The general form of this function is If x is A then y is B. Based on the results of the linguistic variables determining the fuzzy set, the following rules or rule implications can be seen in Table I and set of fire status can be seen in Fig. 11.

TABLE I. FIRE STATUS RULES

Number rule	Temperature	Smoke	Fire	State
1	Normal	Thin	True	Caution
2	Normal	Thin	False	Safe
3	Normal	Medium	True	Danger
4	Normal	Medium	False	Safe
5	Normal	Dense	True	Danger
6	Normal	Dense	False	Danger
7	Warm	Thin	True	Danger
8	Warm	Thin	False	Safe
9	Warm	Medium	True	Danger
10	Warm	Medium	False	Safe
11	Warm	Dense	True	Danger
12	Warm	Dense	False	Danger
13	Hot	Thin	True	Danger
14	Hot	Thin	False	Safe
15	Hot	Medium	True	Danger
16	Hot	Medium	False	Caution
17	Hot	Dense	True	Danger
18	Hot	Dense	False	Danger

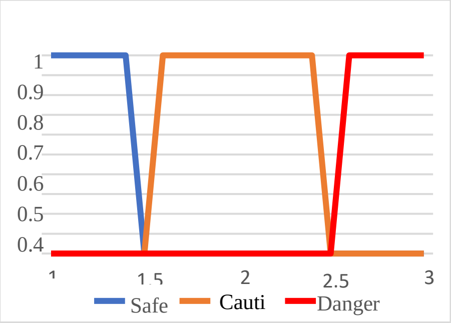


Fig. 11. Set of Fire Status.

At this stage the function used is the minimum calculation, namely by taking the lowest value from the fuzzy set of temperature, smoke, and fire based on the rules that have been made.

$$\alpha i = \mu\ A1\ (X) \cap \mu\ B1\ (X) = \text{MIN}\ \{\mu\ A1\ (X), \mu\ B1\ (X)\} \quad (1)$$

The fire status variable is used to determine the status of the fire status with the Implication function, which consists of safe, alert, and danger with a range as depicted in Fig. 11. That is, safe defuzzification value is 1 – 1.5, Alert 1.5-2.5, and danger 2.5 – 3. Defuzzification

This stage is done by taking the maximum value obtained from taking the minimum value from each rule that has the same rule output.

$$\text{Usf [Xi]} = \text{MAX (Usf [Xi], Ukf [Xi])} \quad (2)$$

- Usf [Xi]*:* value of the i-th order fuzzy solution.
- Ukf [Xi]: value of the i-th order fuzzy solution.

After that, the composition of the maximum value is obtained as follows:

Safe = MAX (rule 2, rule 4, rule 8, rule 10, rule 14)

Caution = MAX (rule 1, rule 3, rule 6, rule 7, rule 16)

Danger = MAX (rule 5, rule 9, rule 11, rule 12, rule 13, rule 15, rule 17, rule 18)

Then Defuzzification is done by entering the maximum value of each output rule into the weighted average formula:

III. Result and Discussion

In Table III, the temperature is stable below 30°C, Smoke is stable in the range of 428ppm to 456ppm and Fire is not detected at all. So based on the rules in Table II, the fire status is Safe.

TABLE II. Test Results with no Fire Conditions

Minute	Temperature (°C)	Smoke (ppm)	Fire	Fire Status
1	28.31	440.63	False	Safe
2	28.19	428.13	False	Safe
3	28.19	428.13	False	Safe
4	28.13	428.13	False	Safe
5	28.13	428.13	False	Safe
6	28.25	428.13	False	Safe
7	28.25	440.63	False	Safe
8	28.19	428.13	False	Safe
9	28.19	456.25	False	Safe
10	28.06	428.13	False	Safe
11	28.13	440.63	False	Safe
12	27.94	428.13	False	Safe
13	27.94	428.13	False	Safe
14	28	440.63	False	Safe
15	28	428.13	False	Safe

In Table III there are three types of fire status that appear. In minutes 1 – 4 there is a fire alert status caused by temperatures exceeding 30°C, smoke in the range of 440ppm to 471ppm, and detected fire. Meanwhile, at minute 5, the status of the fire that occurred was Safe, which was caused by not detecting the fire. This is evidenced by the data at 6 minutes when the fire was detected again, and the fire status returned to Alert. At 9, 13, and 15 minutes there was a Danger fire status, this was caused by a temperature exceeding 45°C, Smoke exceeding 440ppm and a detected fire.

TABLE III. Test Results with 50cm from the Fire

Minute	Temperature (°C)	Smoke (ppm)	Fire	Fire Status
1	32.06	471.88	True	Caution
2	32.13	440.63	True	Caution
3	31.63	471.88	True	Caution
4	35.31	440.63	True	Caution
5	36.56	471.88	True	Safe
6	35.19	440.63	True	Caution
7	37.19	456.25	True	Caution
8	49.25	440.63	True	Caution
9	53.25	609.38	True	Danger
10	48.5	471.88	True	Caution
11	44.63	440.63	True	Caution
12	51.56	440.63	True	Caution
13	47.75	1159.38	True	Danger
14	50.94	428.13	True	Caution
15	53.56	440.63	True	Danger

In Table IV, it can be seen that the farther the distance from the fire source causes the slightly lower temperature detected.

TABLE IV. Test Results with 100cm from the Fire

Minute	Temperature (°C)	Smoke (ppm)	Fire	Fire Status
1	31.63	428.13	False	Safe
2	31.69	428.13	False	Safe
3	31.94	440.63	False	Safe
4	32.13	440.63	False	Safe
5	32.25	440.63	False	Safe
6	32.44	1128.13	False	Safe
7	32.56	440.63	False	Safe
8	32.75	440.63	False	Safe
9	32.88	440.63	False	Safe
10	33.06	440.63	False	Safe
11	33.06	1190.63	False	Caution
12	33.19	471.88	False	Safe
13	33.13	487.5	True	Caution
14	33.25	471.88	False	Safe
15	33.31	487.5	False	Safe

The test results with burning mosquito drugs are shown in Table V.

TABLE V. Test Results with Burning Mosquito Drugs

Minute	Temperature (°C)	Smoke (ppm)	Fire	Fire Status
1	31.63	428.13	False	Safe
2	31.69	428.13	False	Safe
3	31.94	440.63	False	Safe
4	32.13	440.63	False	Safe
5	32.25	440.63	False	Safe
6	32.44	1128.13	False	Safe
7	32.56	440.63	False	Safe
8	32.75	440.63	False	Safe
9	32.88	440.63	False	Safe
10	33.06	440.63	False	Safe
11	33.06	1190.63	False	Caution
12	33.19	471.88	False	Safe
13	33.13	487.5	True	Caution
14	33.25	471.88	False	Safe
15	33.31	487.5	False	Safe

TABLE VI. TEST RESULTS WITH 15 PORTABLE STOVES

Minute	Temperature (°C)	Smoke (ppm)	Fire	Fire Status
1	30.38	428.13	FALSE	Safe
2	30.38	440.63	FALSE	Safe
3	30.31	456.25	FALSE	Safe
4	30.25	440.63	FALSE	Safe
5	30.44	440.63	FALSE	Safe
6	30.38	440.63	FALSE	Safe
7	30.38	1159.38	FALSE	Safe
8	30.44	440.63	FALSE	Safe
9	30.44	440.63	FALSE	Safe
10	30.44	440.63	FALSE	Safe
11	30.44	440.63	FALSE	Safe
12	30.5	428.13	FALSE	Safe
13	30.5	428.13	FALSE	Safe
14	30.63	428.13	FALSE	Safe
15	30.5	428.13	FALSE	Safe

The test results are shown in Table VI, it can be concluded that the false alarm detection accuracy is 100%.

Fuzzy logic testing is done by comparing the fuzzy output value on a series of systems with fuzzy output values found in web. Test performed 10 times is shown in Table VII.

TABLE VII. TEST RESULTS WITH 10 PORTABLE STOVES

Input			Output		Difference	Error
Temperature (°C)	Fire	Smoke (ppm)	Arduino	Web		
23.4	48.8	26.5	2	1.28	0.72	5.62%
29.25	774	1.31	2	2	0	0%
24	50	20	2	1.4	0.6	4.28%
25	25	280	4	4	0	0%
35	25	300	5	5	0	0%
15	500	25	1	1	0	0%
25	200	100	2.73	2.73	0	0%
25	350	200	3	3	0	0%
10	800	20	1	1	0	0%
15	200	500	4	4	0	0%
Average					0.13	0.99%

IV. CONCLUSION

Based on the web-based and WhatsApp-based Forest fire monitoring dashboard detection application that has been done, it can be concluded that: applications can be used by forest halls and forest rangers, the application is implemented by utilizing Arduino along with Arduino sensors to get real time data, the application implements Sugeno Fuzzy Inference for classify fire status and detect false alarm based on data received from Arduino.

Based on the results of tool design and testing, it can be concluded that the performance of the tool as an early fire detector, also functions as a monitoring system for room conditions against potential fires, is able to represent the results of reading all multisensory data and can classify the condition of the room being monitored properly. According to the input data received by the hardware and software. In testing the fuzzy program, the fuzzy output between Arduino and the fuzzy on the monitoring dashboard has a small difference of 0.99%. This shows that fuzzy logic control on multisensory data reading has a high level of accuracy.

ACKNOWLEDGMENT

This work is supported by Research and Technology Transfer Office, Bina Nusantara University as a part of Bina Nusantara University's International Research Grant (PIB 2021) with contract number: 017/VR.RTT/III/2021.

REFERENCES

[1] S. Sudhakar, V. Vijayakumar, C.S. Kumar et al., Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires, Computer Communications (2019), doi: https://doi.org/10.1016/j.comcom.2019.10.007.

[2] Pryanka, A. 14 November (2018). Potensi Pasar Internet of Things di Indonesia Capai Rp 444 T. (Online). Diakses 8 Juli 2021 dari https://republika.co.id/berita/pi6068383/potensi-pasar-eminternet-of-thingsem-diindonesia-capai-rp-444-t..

[3] Purnomo, H.; Shantiko, B.; Sitorus, S.; Gunawan, H.; Achdiawan, R.; Kartodihardjo, H.; Dewayani, A.A. Fire economy and actor network of forest and land fires in Indonesia. Forest Policy and Economics 78: 21-31. 2017.

[4] Evizal Abdul Kadir1, Sri Listia Rosa, Rizdqi Akbar Ramadhan. Detection of Forest Fire Used Multi Sensors System for Peatland Area in Riau Province. International Conference on Industrial, Mechanical, Electrical and Chemical Engineering (ICEMECE) 2019.

[5] Josue Toledo Castro, Pino Caballero Gil, et al. Forest Fire Prevention, Detection and Fighting Based on Fuzzy Logic and Wireless Sensor Network. Hindawi Volume 2018. https://doi.org/10.1155/2018/1639715.

[6] T. Saikumar, P. Sriramya. Iot Enabled Forest Fire Detection and Altering the Authoritie. International Journal of Recent Technology and Engineering (IJRTE) Volume 7, Issue 654, April 2019.

[7] Dernoncourt, F. (2013). Introduction to fuzzy logic. Cambridge: Massachusetts Institute of Technology.

[8] Masayu Annisah, Nyayu Latifah Husni, Tresna Dewi, RM Aprillia Rachmawati. Peat Land Fire Monitoring System Using Fuzzy Logic Algorithm. Computer Engineering and Applications Vol. 8, No. 3, October 2019.

[9] Dharmawan, A., Budiman, A., Wijaya, A., Margono, B. A., Martinus, D., Ridha, D. M., . . . Rusolono, T. (2015). National Forest Reference Emissions Level For Redd+.

[10] Chang, S. S. (2016). Go Web Programming. Shelter Island: Manning Publications.

[11] Levinson, D., & Belton, T. (2017). Build your first Web app : learn to build Web applications from scratch. New York: Sterling Publishing.

[12] McRoberts, M. (2013). Beginning Arduino. New York: Apress.

[13] Sun, S. J., Zhang, Z. Q., & Han, H. (2017). Study on the Method of Power Network Operation Early Warning Using Remote Sensor Monitoring and Locating of Satellites Based on Power Grid GIS. Electric Power, 50(4), 181-184. http://www.chinapower.org/CN/10.11930/j.issn.1004-9649.2017.04.181.04.

[14] Zhang, C., & Yang, H. Q. (2017). Study on Fire Positioning Scheme and Fire Alarm Warning Based on Infrared Technology. Measurement & Control Technology, 36(7), 33-37. http://www.cnki.net/kcms/doi/10.19708/j.ckjs.2017.07.009.html.

[15] Li, Y. J., Zheng, W., Chen, J., & Liu, C. (2017). Fire Monitoring and Application Based on Meteorological Satellite. Aerospace Shanghai, 34(4), 62-72. http://www.cnki.net/kcms/doi/10.19328/j.cnki.1006-1630.2017.04.008.html.

[16] Sasmoko, D., & Mahendra, A. (2017). Rancang Bangun Sistem Pendeteksi Kebakaran Berbasis IOT dan SMS Gateway Menggunakan Arduino. Jurnal Simetris, Vol 8, no. 2,(pp.469-476).

[17] MS Anggreainy, A Wulandari, AM Illyasu. Diagnosing COVID-19 symptoms using fuzzy logic. 2021 5th International Conference on Informatics and Computational Science.

[18] U. P. Bisba, E. Mulyana, M. A. Ramdhani and M. Irfan, "The Implementation of The Fuzzy Sugeno Algorithm On an IoT-Based Temperature and Humidity Monitoring System," 2019 IEEE 5th International Conference on Wireless and Telematics (ICWT), 2019, pp. 1-6, doi: 10.1109/ICWT47785.2019.8978217.

Implementation of Gamification in Mathematics m-Learning Application to Creating Student Engagement

Sufa Atin, Raihan Abdan Syakuran, Irawan Afrianto
Informatics Engineering Department
Universitas Komputer Indonesia
Bandung, Indonesia

***Abstract*—Mathematics is one of the main subjects in school. In some schools, the learning methods used are still using conventional methods, namely lectures and exercises. The main difficulty in learning mathematics is how to make the material presented more interesting so that it does not make students bored and easy to understand the material. The use of an attitude of interest in games that knows no age and the various advantages of games gives rise to a combination of learning mechanisms called gamification. Gamification is the process of applying game mechanics to non-game activities to increase user interactivity. Gamification in the m-learning mathematics application was developed using the Attention, Relevance, Confidence, and Satisfaction (ARCS) learning model and the octalysis framework gamification method. Gamification in this mathematics m-learning application applies a game strategy using a system of levels, missions, challenges, points, progress bars, leader boards, and badges. The results of this study indicate that this application can be used as an alternative medium for learning mathematics and student engagement with the result that gamification applied to the m-learning mathematics application can increase student interest by 35%, increase student motivation by 33%, and improve understanding 42% of students towards learning mathematics.**

Keywords*—*Gamification; m-learning; mathematics; attention; relevance; confidence; and satisfaction (ARCS) model; octalysis framework; student engagement

I. Introduction

Mathematics is one of the disciplines that is studied at all school levels, given from elementary students to higher levels [1]. Mathematics is considered a difficult subject because the characteristics of mathematics are abstract, logical, systematic, and full of confusing symbols and formulas [2]. This is because mathematics learning is still conventional, which causes teachers to have difficulty developing the material contained in the book because of the large number of materials to be taught, while from the student aspect, students' lack of interest in learning mathematics, understanding of concepts that are not mature, enthusiasm for learning is high. less and students are not motivated when learning mathematics, and many students view mathematics as a difficult and boring subject[3].

The development of ICT has helped a lot in the field of mathematics education [4]. Computer-assisted mathematics learning applications have been widely developed as alternative learning media [5][6]. Mobile application development (m-learning) is one of the media that is currently widely used in the development of learning applications other than e-learning [7]. This is supported by easy access [8], a more flexible learning process [9], as well as its ability to provide interactive and communicative mechanisms [10] to be one of the causes of the development of e-learning learning models toward m-learning [11].

Gamification is the process of using in-game mechanisms or rules for non-game activities to increase user interactivity [12]. Gamification offers application design that embeds game elements so that it has more appeal to application content because game concepts are known to be fun and easy to understand. [13]. This is because the gamification application model must have the characteristics of attracting the attention of students to use it, able to build student motivation and competence, build student confidence and be interactive to build reasoning and mindset when completing tasks and satisfaction, namely students feel happy when they complete their assignments. [14][15]. To develop a gamification mechanism in a learning application, a reference model is needed [16]. The ARCS learning model unites several forms of student attitudes, namely attention, relevance, confidence, and satisfaction [17], where this model is designed to encourage student learning motivation by prioritizing attention to students [18], as well as applications that adapt the subject matter to the student's learning experience, which can create student confidence and create a sense of satisfaction for the student to study harder [19]. In order to develop the ARCS model, an octalysis gamification framework is needed where the gamification concept is designed by analyzing eight aspects of game psychology [20], which is a gamification concept with an emphasis on human-focused design [21], to optimize human motivation in a system, as opposed to function-focused design [22].

From the previous literature, the implementation of mobile applications, gamification techniques, and the ARCS learning model is still being developed partially, there has been no research that has tried to combine these three elements in a unified system. Therefore, the objective of this research is to develop a model and application of gamification-based mathematics m-learning by integrating the ARSC model and octalysis framework. The implementation of the application is carried out to test and prove whether the concepts and applications that have been developed previously can increase student involvement in mathematics.

II. Literature Review

A. Gamification

Gamification is a product, a way of thinking, a process, an experience, a way of design, and a system that is involved, which uses game elements to solve non-game problems. [23][24]. In the world of education, it can also be said that gamification is a process of changing existing activities or studying activities and making learning content like a game [25]. Gamification is using game mechanics to provide practical solutions by building specific group engagement [26]. In more detail [27], defines gamification as a concept that uses game-based mechanics, aesthetics, and game thinking to engage people, motivate action, promote learning and solve problems. In addition, gamification also provides additional motivation to ensure that students participate in complete learning activities [28]. The concept of gamification in an educational environment has the aim of maximizing students' learning comfort, thereby inspiring and motivating them to continue their learning process [29]. The Gamification Model proposed in 2015 (Fig. 1), consists of elements Mechanics which refers to the elements that comprise gamification, Measurement which refers to how progress in gamification is evaluated. Behavior - refers to the desired actions that players will cultivate as a result of playing the game, and Rewards - refers to the types of incentives given to players for fulfilling a requirement or task in the game [30].

B. ARCS Learning Model

Attention, Relevance, Confidence, and Satisfaction (ARCS) is a learning model developed as an alternative that can be used by teachers to motivate student learning by carrying out learning activities well [31]. This learning model contains four components that are an integral part of learning activities, namely (1) generating and maintaining student attention during the learning process (Attention), (2) providing subject matter relevant to students (Relevance), and (3) providing truest self to students (Confidence), and (4) foster student satisfaction with the learning process (Satisfaction) [32]. Within the teaching framework, the motivational analysis must be an ongoing process, to ensure that gamification matches the motivational factors as learning takes place [33]. The ARCS model has provided several specific steps to examine the relevant motivational features in the use of various media in the learning process [34][35]. Fig. 2 shows the ARCS Model applied to an educational game-based learning application.

C. Octalysis Gamification Framework

Octalysis is a gamification framework developed by Yu-kai Chou in 2015. The Octalysis method has two levels, where the first level is an analysis of game elements from the Octalysis framework, while the second level is the application of game elements in four phases that have been provided. Octalysis is based on a gamified framework designed using eight core drives [36]. Fig. 3 shows the gamification framework using the octalysis method at the first level, which consists of eight core drives in the octalysis framework, namely: (1) Epic Meaning and Calling: in this drive, someone is convinced that they are doing something great or feel that they are the chosen ones for doing something, (2) Development and Accomplishment: internal drive to make progress, develop skills and overcome challenges, (3) Empowerment of Creativity and Feedback: users engage in creative processes such as trying different combinations, (4) Ownership and Possession: these drives make users motivated because they feel like they have something, (5) Social Influence and Relatedness: drives that come from the environment, (6) Scarcity and Impatience: the drive to get something for not having it, (7) Unpredictability and Curiosity: the drive to know what will happen next, and (8) Loss and Avoidance: this drive is based on avoiding something negative happening [37].

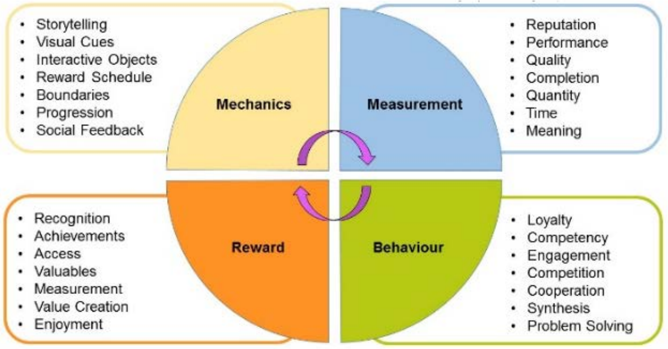
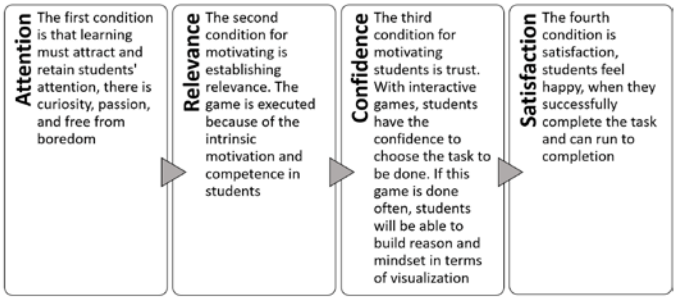


Fig. 1. Gamification Model of Learning. Available from: http://ivantehrunningman.blogspot.com/2015/04/gamification-oflearning.html.



Attention
The first condition is that learning must attract and retain students' attention, there is curiosity, passion, and free from boredom

Relevance
The second condition for motivating is establishing relevance. The game is executed because of the intrinsic motivation and competence in students

Confidence
The third condition for motivating students is trust. With interactive games, students have the confidence to choose the task to be done. If this game is done often, students will be able to build reason and mindset in terms of visualization

Satisfaction
The fourth condition is satisfaction, students feel happy, when they successfully complete the task and can run to completion

Fig. 2. ARCS Model in Educational Game Learning Media.

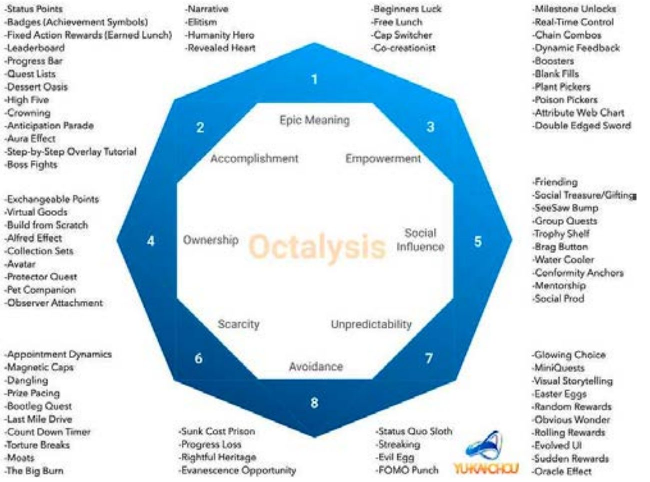


Fig. 3. Level I Octalysis Gamification Framework.

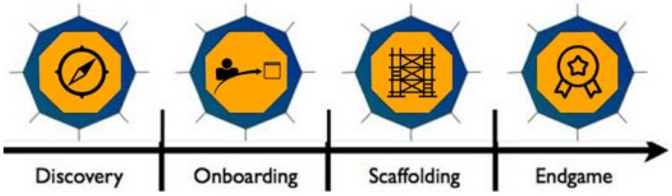


Fig. 4. Level II: Octalysis Level II Gamification Framework.

After applying the game elements that will be used, the elements are arranged into four phases which are part of the octalysis framework, level II. Fig. 4 The following shows the four phases of the second level octalysis framework [38]: (1) Discovery phase which is the initial stage where new users enter the application system and introduction of the application, (2) The onboarding phase is the phase where users begin to get to know the flow and application rules, (3) The scaffolding phase is the phase where users start using the application after getting to know the flow and the main mission of the application, and (4) The endgame phase aims to keep players using the application after the goal of the application has been achieved.

D. Multimedia Development Life Cycle (MDLC)

In the field of learning, the use of multimedia has been widely used, starting from the use of text, images, animation, video, and audio, to motivate students to like teaching materials. Fig. 5 shows the software development method used, namely the Multimedia Development Life Cycle (MDLC) introduced by Luther [39] and developed by Sutopo [40].

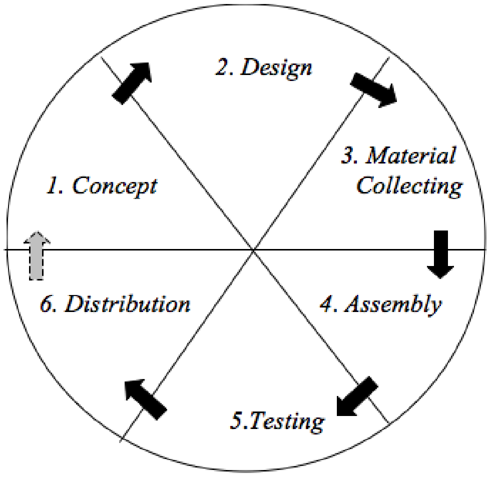


Fig. 5. Multimedia Development Life Cycle.

The concept is the activity to determine the purpose and who are the users of the program [41], Design is the stage of making specifications regarding the program architecture, style, appearance, and material/material requirements for the program [42], Material collecting is the activity of collecting materials following the needs being worked on [43], Assembly is the activity of making all multimedia objects or materials. Application development is based on the design stage, such as storyboards, flowcharts, and/or navigation structures [44], Testing is carried out after completing the assembly activity by running the application program and seeing whether there are errors or not [45], and distribution where the application will be stored in a storage medium. This activity can also be called the evaluation part to develop a finished product to make it better. The results of this evaluation can be used as input for the concept stage of the next product [46].

III. Research Method

This research was conducted using a quantitative descriptive research method [47] namely a research method that provides an objective description of an existing problem analytically and measurably [48].

Fig. 6 shows the five stages of research design carried out in this study, namely problem formulation, data collection, software development using MDLC, and conveying the results.

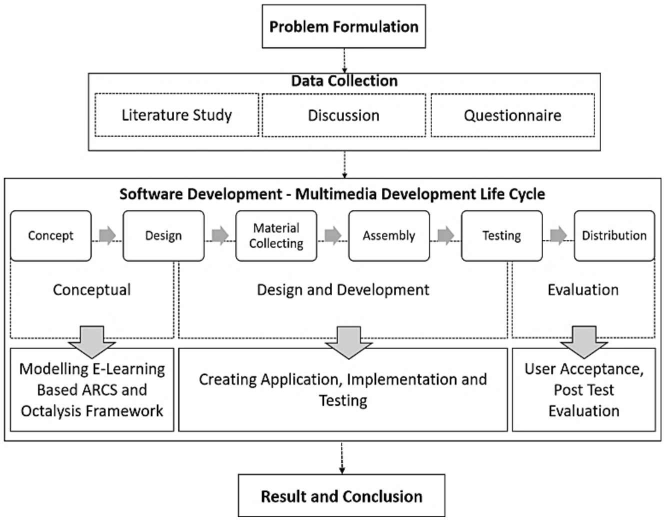

Problem Formulation
Data Collection
Literature Study
Discussion
Questionnaire
Software Development - Multimedia Development Life Cycle
Concept
Design
Material Collecting
Assembly
Testing
Distribution
Conceptual
Design and Development
Evaluation
Modelling E-Learning Based ARCS and Octalysis Framework
Creating Application, Implementation and Testing
User Acceptance, Post Test Evaluation
Result and Conclusion

Fig. 6. Research Design.

The first stage is problem formulation which is an attempt to uncover various things related to the problem to be answered or solved. In this study, the formulation of the problem was carried out by identifying the existing mathematics learning problems, what are their shortcomings, and the solutions offered for their completion. This is the aim of the research conducted.

The second stage is the collection of research data utilizing literature studies, discussions, and questionnaires. The literature study was carried out by collecting previous studies related to the ARCS learning model, the theory and implementation of gamification, and the mechanism of MDLC software development. In addition, discussions with the school, namely teachers and students in mathematics, were carried out to collect data on learning materials, get input and describe the desired solution. Meanwhile, the use of questionnaires was carried out for scientific measurement of the results obtained from this research.

The fourth stage is the development of m-learning mathematics software. In this research, software development is carried out using the MDLC method, which starts from the conceptual section which contains the system model to be developed, the users, and its functions in it. Followed by the design and development section, where the system design is carried out, making the assets in the system developed, coding the system, and conducting initial testing. The last part at this stage is an evaluation by testing the system both functionally and in terms of acceptance of its users.

The fifth stage is to summarize the results that have been obtained, provide conclusions, state the limitations in the system built, and direct the development of further research.

IV. Result and Discussion

A. Analysis of Current Learning Conditions

Based on the discussion conducted by the researcher using interviews and field observations (Fig. 7), there are two things found in the learning process of mathematics subjects, namely that the learning process is still done conventionally. The teacher provides learning materials through books, blackboards, and written exams. Meanwhile, students receive learning by listening and doing the tests given by the teacher. On the other hand, there are no alternative tools that can be used by students in the learning process, limiting students to be able to learn and understand mathematics subject matter.

Based on these problems, a solution was developed to build m-learning applications for mathematics subjects that can be used by students to learn from anywhere, with more interactive and interesting content in the form of gamification in mathematics subjects.

B. Conceptualizing Mathematics M-Learning Gamification

Fig. 8 shows the conceptualization of the application carried out to describe what things will be done by the system along with its methods and work functions. Application development starts from user needs, implementation of the ARCS model, and gamification in the application, followed by application development to achieve the expected goals.

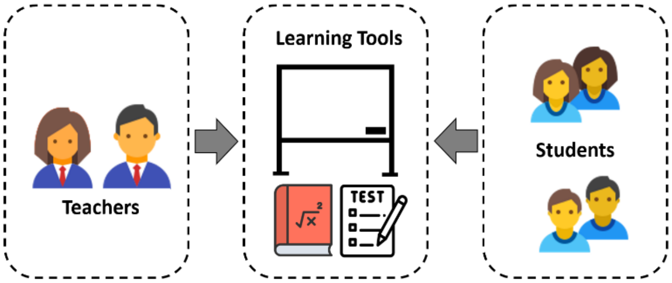


Fig. 7. Existing Learning Activities.

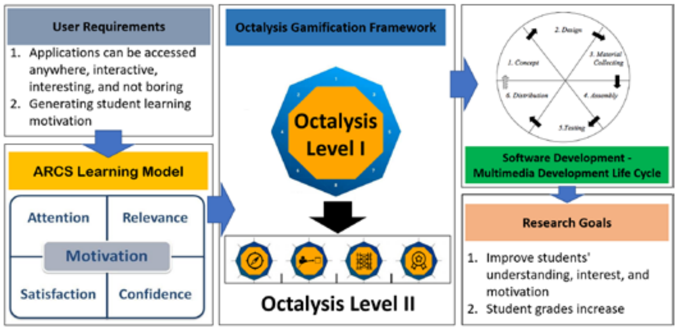

User Requirements
1. Applications can be accessed anywhere, interactive, interesting, and not boring
2. Generating student learning motivation
ARCS Learning Model
Attention
Relevance
Motivation
Satisfaction
Confidence
Octalysis Gamification Framework
Octalysis Level I
Octalysis Level II
2. Design
3. Material Collecting
1. Concept
4. Assembly
6. Distribution
5. Testing
Software Development - Multimedia Development Life Cycle
Research Goals
1. Improve students' understanding, interest, and motivation
2. Student grades increase

Fig. 8. Conceptual Development of Mathematics Gamification Applications.

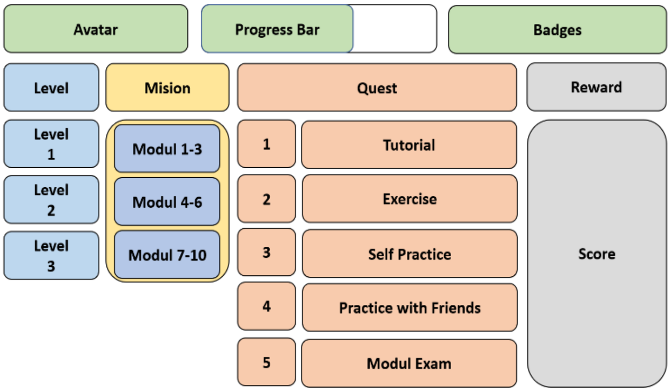


Fig. 9. Mathematics m-learning Gamification Model.

Gamification modeling in the m-learning mathematics application (Fig. 9) was developed by providing educational game effects in the form of game levels related to the learning module for mathematics materials, quests which are challenges that must be completed by students, rewards in the form of points and badges, and avatar that can be customized by students.

C. Octalysis Gamification Analysis

The octalysis method has eight core elements, each of which has a game technique that can be chosen to be implemented. Mathematics learning applications are built using the concept of gamification with the octalysis framework method. This method begins with the level 1 stage by analyzing the elements that will be applied, and then enters the second stage, namely, level 2 making gameplay.

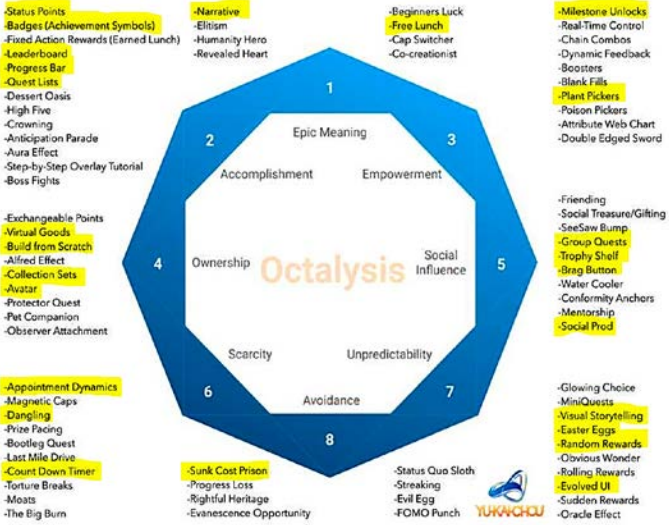


Fig. 10. Octalysis Gamification Elements are used.

Fig. 10 shows that there are 25 elements of the octalysis framework used in the development of this gamification application.

1) Elements of Epic Meaning and Calling, the application convinces the user that he is doing something bigger than himself or that he is "chosen" to do something. Game techniques used in the application to be built include:

a) Narrative, the application will start with a narrative or story that gives an idea of why the user should play the game or use this application.

b) Free Lunch, to attract users there is a limited item for 20 users at the start of the game.

2) Development and Accomplishment elements are used to encourage users to make progress, develop and overcome existing challenges or tasks. Game techniques used in the application to be built include:

a) Progress Bars, there are 2 types of progress bars in the application. The first progress bar is useful for showing the progress of the usage adventure, this progress bar illustrates how far the user has completed the existing material, for example, the user has completed 5 of the 6 chapters that are displayed using the bar. The second progress bar is used to show the user's ability or results in completing each material.

b) Achievement Symbols are used in the form of badges or badges. Every completed action or task will get feedback in the form of a badge. Status Points: each user has status points as a measuring tool or assessment in carrying out existing tasks.

c) The leaderboard is used as a facility to compare the achievements of one player with other players. The use of the leaderboard also aims to motivate players to always feel challenged to be the best.

d) Quest List, is used to displaying what tasks or challenges the user has to do in the application.

3) Empowerment is an element that is made so that users can be as creative as possible in solving problems. The following are the game techniques used in the application to be built:

a) Milestone Unlock: Correlating with the quest list, the existing materials, exercises, and exams cannot be accessed directly by the user. Availability depends on the related task, if the related task has been completed; it means that the user has met the requirements to access the material in question.

b) Plant Picker: Each user has different abilities; they can freely determine the reward they choose.

4) Ownership and Possession are elements where users are motivated because they feel like they own something. When a player feels ownership, he innately wants to make what he has better and have more of it. The following are the game techniques used in the application to be built:

a) Avatar: Each user has their avatar.

b) Virtual Goods: Correlating with the avatar game element, where the existing avatar is created based on the user's wishes, the user can choose the hair, clothes, and eyes for his avatar.

c) Build-From-Scratch: User avatars can be changed by purchasing items in the shop and using limited items from quest/task rewards.

d) Collection Sets: Users can provide feedback in the form of likes to their friends for group activities or assignments.

5) Elements of Social Influence and Relatedness are used to encourage users about guidance, competition, jealousy, group seeking, social possessions, and friendship so that users continue to use existing applications. The following are the game techniques used in the application to be built:

a) Trophy Shelves: Trophy Shelves are implemented in the leaderboard and profile where each user can see other users' searches.

b) Group Quest: There are several materials where training sessions must be carried out together, with this game technique users can interact with each other.

c) Social Prods: Implemented in the form of “likes” after doing Group Quests as a form of appreciation for fellow users.

6) The element of Scarcity and Impatience is the drive that motivates us simply because we cannot have something immediately, or because there is great difficulty in getting it. The following are the game techniques used in the application to be built:

a) Dangling and Anchored Juxtaposition: Energy usage limits the number of tasks taken per day.

b) Appointment Dynamics: This game technique is implemented using push notifications, where users will be given a message every week to use this application.

7) Unpredictability and Curiosity elements are user urges to find out what will happen next. The following is the game technique used in the application to be built:

a) Mystery Boxes/Random Rewards: There are mystery box items as rewards in several missions.

b) Easter Eggs/Sudden Rewards: There are limited rewards on some missions if the user reaches certain conditions.

c) Visual Storytelling: Dealing with Narrative elements where these elements are used as a visualization of the existing narrative.

d) Evolved UI: The existing UI will evolve according to the user level.

8) The Loss and Avoidance element motivates the user through the fear of losing something. For example, if our mission gets a low star rating, the user's points are reduced. The following are the game techniques used in the application to be built:

a) The Sunk Cost Prison: Due to dangling and anchored juxtaposition, the user's energy will be reduced if he exits the application while carrying out certain processes, for example in the training process.

b) Countdown Timers: related to the sunk cost prison and Dangling and Anchored Juxtaposition after the energy runs out the user cannot perform activities for a certain time.

The game elements that have been obtained at the level I are applied to four stages of gamification level II Octalysis namely discovery, onboarding, scaffolding, and endgame which can be seen in Table I.

TABLE I. ELEMENT GAMIFIKASI OCTALYSIS LEVEL II

Phase	Description
discovery	The discovery phase is the initial stage where a new user enters the application system and introduces the application. In this phase, the application uses several technical games, namely Narrative, Visual Storytelling, Free Lunch, and Avatar. Users are first presented with a story that describes why they should use this application and complete the missions according to the Narrative and Visual Storytelling game techniques. After that, the user creates his avatar as a character who will carry out the existing mission. The introduction ends with the award of points and gold as a form of Free Lunch.
onboarding	The Onboarding phase is the phase where the user gets to know the flow and rules of the application. The implementations in it are Progress Bars, Badges, Status Points, and Evolved UI. Students enter the main page in the form of a UI that displays the selected character. Then, students can see Status Points in the form of Points and Energy which are the value handles while using the application. Students can also see various assignments and their rewards on the Quest List page.
scaffolding	This phase is the phase where users start using the application after getting to know the flow and main mission of the application. Users perform activities according to the Quest List to achieve the goals of the application. Users can see the ranking of learning achievements through the leaderboard of accumulated points from completed tasks, besides that users can see the badges earned on the profile page (Trophy Shelves). At this stage, the user is presented with a challenge in which the user evaluates in the form of doing exercises and exams using the Fisher-Yates Shuffle algorithm to reduce the level of cheating. There are joint exercises as the implementation of the Group Quest and each user can give appreciation to his opponent in the form of likes as the implementation of the Social Prod. The results of practice and exams are visualized grades with stars. Each completed Quest earns rewards. Users then have activities to get Badges, according to existing conditions. Here implements Social Treasures and Collection Sets, and there is also a Milestone Unlock where this will unlock badges, as well as Easter Eggs. With the Easter Eggs mechanism, learning can be determined based on the user's decision, this is in line with the plant picker component game. The use of the application is limited according to the amount of energy remaining (The Sunk Cost Prison and Dangling and Anchored Juxtaposition), if the energy runs out the user must wait for a certain time (Countdown timer). Previously created avatars can be changed and customized via items purchased on the implementation avatar pages of Virtual Goods and Build-From-Scratch.
endgame	The last phase is the Endgame phase. This phase aims to keep players using the application after the goals of the application have been achieved. The implementation of this phase uses Appointment Dynamics where users will get notifications periodically.

D. ARCS Model and Octalysis Gamification Mapping

This stage is carried out by mapping the ARCS learning model based on the Attention, Relevance, Confidence, and Satisfaction categories, as well as the octalysis method in application development.

Table II shows the mapping mechanism of the ARCS learning model in the Attention category and the octalysis method.

TABLE II. MAPPING MODEL ARCS (ATTENTION) AND OCTALYSIS METHOD

ARCS Model - Attention			
Sub Category	***Description Category***	***Element Octalysis***	***Description***
Perceptual arousal	Media must have things that can attract users	*Narrative, Visual Storytelling*	The use of narration and visual storytelling to explain why they should use the app and learn the lessons.
		Avatar	Users have their avatar where the appearance of the avatar can change according to student achievements.
	Perception of stimulation through surprise	*Free lunch*	Students who access the app first get special rewards
		Easter Eggs (Sudden Reward)	There is a limited reward if the user reaches certain conditions
		Appointment Dynamic	Students are always reminded every week to study through notifications
	Perception of design through uncertainty	*Mystery Boxes (Random Rewards)*	Students can get random items from the mystery box
Variability	Interesting presentation of material	*Narrative, Tutorial*	Video media is used so that students do not get bored easily.
Inquiry arousal	Students can study the material independently	*Narrative, Tutorial*	With the narration in the learning video, students can learn independently
	Students can determine the learning process	*Plant picker*	The learning flow of each student is different, students can determine their process

Table III shows the mapping mechanism of the ARCS learning model in the Relevance category and the octalysis method.

Table IV shows the mapping mechanism of the ARCS learning model in the Confidence category and the octalysis method.

TABLE III. MAPPING MODEL ARCS (RELEVANCE) AND OCTALYSIS METHOD

ARCS Model - Relevance			
Sub Category	***Description Category***	***Element Octalysis***	***Description***
Goal Orientation	Explanation of learning objectives	*Narrative*	Before students can see the learning video, the learning syllabus is presented first.
Motive Matching	Explanation of the benefits of learning	*Narrative*	Before students can see the learning video, the learning syllabus is presented first.
Familiarity	Learning adaptation to students	*Narrative*	Examples of problems in learning adapted to the lives of students today

TABLE IV. MAPPING MODEL ARCS (CONFIDENCE) AND OCTALYSIS METHOD

ARCS Model - Confidence			
Sub Category	***Description Category***	***Element Octalysis***	***Description***
Learning Requirements	Learning requirements	*Milestone Unlock*	Students must meet certain requirements to carry out a lesson
		Quest List	Students can see the criteria for taking existing subjects
Success Opportunities	Provides many, varied, and challenging experiences that enhance learning success	*Group Quest*	The available exercises are divided into 2 types, namely individual training and group training (multiplayer)
Personal Control	There is feedback on student learning outcomes	*Status, Progress Bar*	Learning results can be seen through the number of stars and progress bar
		Rewards/ Achievement Reward	Every time you complete a quest, students get rewards
	Student responsibility for learning	*Dangling and Anchored Juxtaposition*	Students cannot arbitrarily carry out continuous learning, learning activities are limited by energy so students must be able to take advantage of their time
		The Sunk Cost Prison	Related to Dangling and Anchored Juxtaposition, students must be responsible if they cannot complete the learning process or leave in the middle of the process.
		Countdown Timers	Students cannot carry out the learning process, there is a time back, the rewards of the student's irresponsibility during the learning process

Table V shows the mapping mechanism of the ARCS learning model in the Satisfaction category and the octalysis method.

E. Analysis System Architecture

M-learning application development has three subsystem architectures in it. The first sub-system is the frontend application which functions as a gamification application used by students, a backend application used by teachers to manage the system and monitor student learning outcomes, and the internet which is used for the exchange and storage of application data. The architecture of the mathematical m-learning application can be seen in Fig. 11.

Functionally, the frontend application is an application that has gamification features. This application is used by students to learn math subjects.

TABLE V. MAPPING MODEL ARCS (SATISFACTION) AND OCTALYSIS METHOD

ARCS Model - Satisfaction			
Sub Category	***Description Category***	***Element Octalysis***	***Description***
Intrinsic reinforcement	Internal satisfaction that can motivate students	*Badges (Achievement Symbol)*	Students get rewards in the form of badges as a visualization of their achievements
		Trophy Shelves	All learning achievements can be seen through the profile page
		Virtual Goods, Build From Scratch	*Existing avatars can be changed and created according to the user's wishes*
		Evolved UI	The existing UI changes according to user achievements
Extrinsic Rewards	External achievements that can motivate students	*Status, Leaderboard*	There is a student leaderboard to motivate which is obtained from accumulated points (status) obtained from rewards
Equity	Students can get learning feedback from other students	*Social Prod, Collection Sets*	Students can give feedback "like" after doing the exercise together

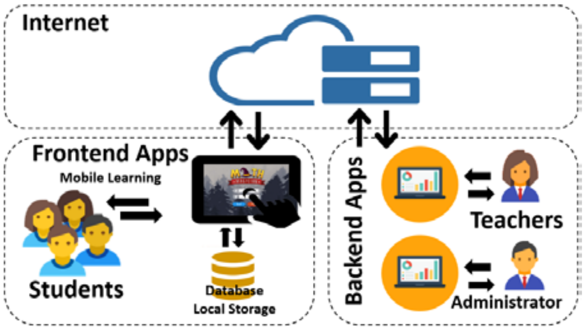


Fig. 11. Mathematics m-learning Application Architecture.

Fig. 12 shows the functionalities developed in the mathematics m-learning application, where there are login functions, viewing materials, doing exercises, taking exams, viewing quests, viewing profiles, viewing the leaderboard, and creating avatars.

F. Gamification Logic Design

To develop the functionalities described in the use case diagram, a logical design is made for each function that shows the learning flow that can be carried out in the application of m-learning mathematics. The flow can be seen in Fig. 13 which starts from accessing the module to its completion, calculating scores and badges as an award for student achievement of the material that has been completed.

1) Tutorial design process: Learning materials (tutorials) are studied using a linear tutorial model and branched tutorials depending on the actions taken by students according to the application of game elements in the previous analysis. The linear tutorial causes students can't able to choose freely to access the material they want to learn, and must follow the sequence according to the order of the existing material. Meanwhile, in the branching tutorial, students can take other materials if they have special items through random rewards. The tutorial flow can be seen in Fig. 14.

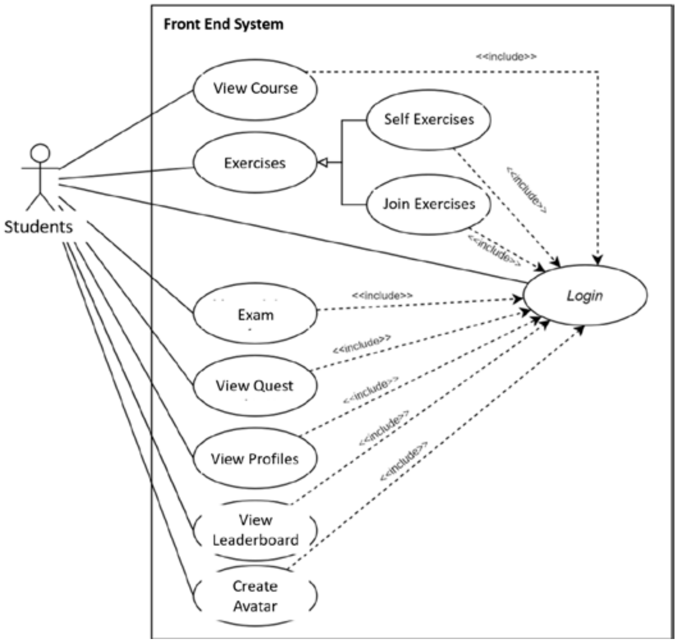


Fig. 12. M-learning Application Frontend use-case Diagram.

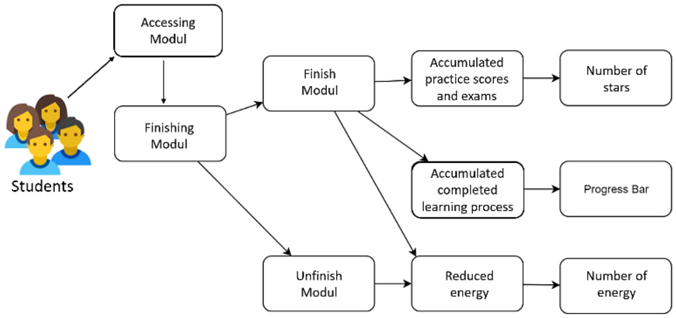


Fig. 13. Learning Flow in Gamification Applications.

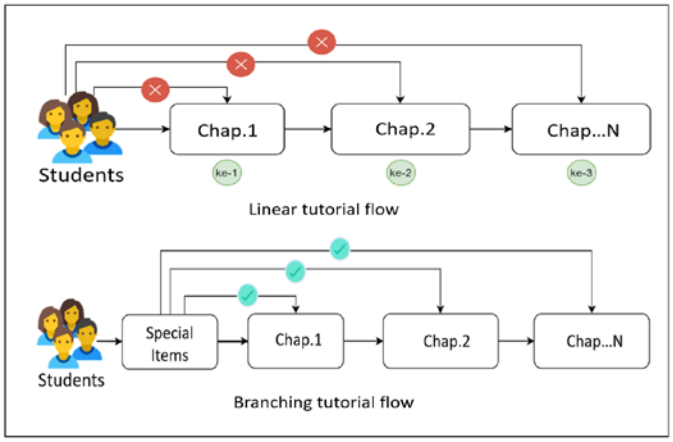


Fig. 14. Tutorial Flow.

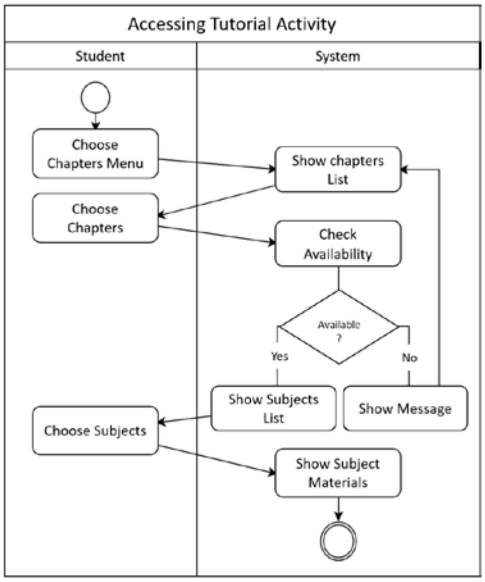


Fig. 15. Accessing Tutorial Activity Diagram.

Fig. 15 is an activity diagram that shows how students access the tutorial functions in the system. The tutorial consists of chapters and subjects of mathematics lessons that students can choose from.

2) Exercise and exam design process: The evaluation process is carried out using two methods, namely practice questions and exams. Students can do 2 practice modes, namely self-practice (Fig. 16), joint practice or multiplayer (Fig. 17), and independent exam (Fig. 18). At the time of practice, the number of questions that came out was 10 questions and 5 questions for joint practice. Meanwhile, at the time of the exam, the questions that came out were 15 questions, in which the questions were in the form of multiple choices.

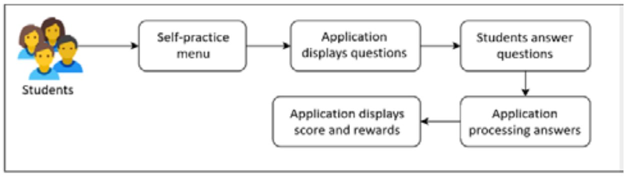

Students
Self-practice menu
Application displays questions
Students answer questions
Application processing answers
Application displays score and rewards

Fig. 16. Self-practice Evaluation Flow.

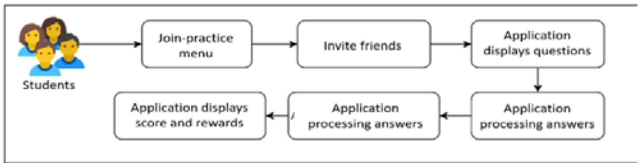

Students
Join-practice menu
Invite friends
Application displays questions
Application processing answers
Application processing answers
Application displays score and rewards

Fig. 17. Joint Practice Evaluation Flow.

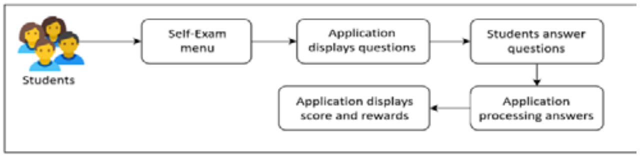

Students
Self-Exam menu
Application displays questions
Students answer questions
Application processing answers
Application displays score and rewards

Fig. 18. Self-examination Evaluation Flow.

There is a special feedback feature after doing joint exercises, where students can give likes to other students after the exercise is finished. This is to give appreciation and socialization between students. The feedback flow can be seen in Fig. 19.

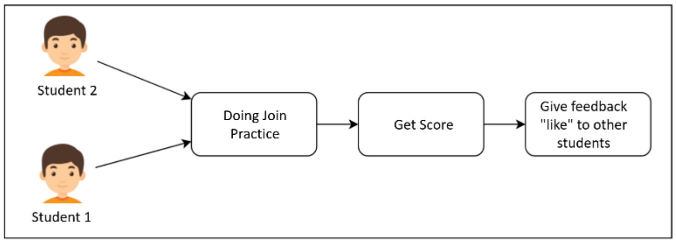


Fig. 19. Joint Practice Feedback Flow.

Fig. 20 shows an activity diagram of the exercise and exam functions that can be performed by students. The function of exercise and exam is an evaluation carried out by the system when a student has completed a chapter of learning in the system.

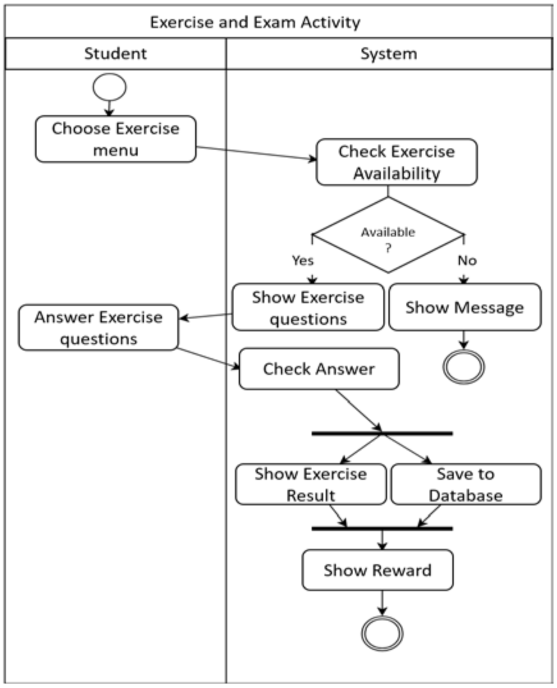


Fig. 20. Exercise and Exam Activity Diagram.

3) Quest flow design: Not all existing processes can be accessed freely by users; certain requirements must be completed first. The following is an analysis of the quest flow in the application that was built, which can be seen in Fig. 21.

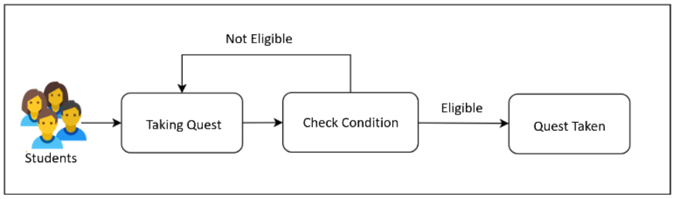


Fig. 21. Quest Selection Flow.

4) Rewards flow design: The form of rewards is adjusted to the status used in existing applications; rewards can be in the form of points or badges for students who have completed an activity. The flow of rewards can be seen in Fig. 22.

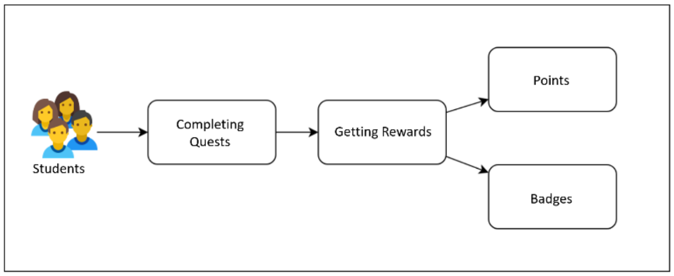


Fig. 22. The Flow of Rewards.

5) Leaderboard flow design: The leaderboard is obtained from the accumulation of student learning outcomes, obtained from the completion of tutorials, completion of exercises and exams as well as other activities. Fig. 23 shows the workflow of the leaderboard creation activity as well as an activity diagram to access the leaderboard contained in the system.

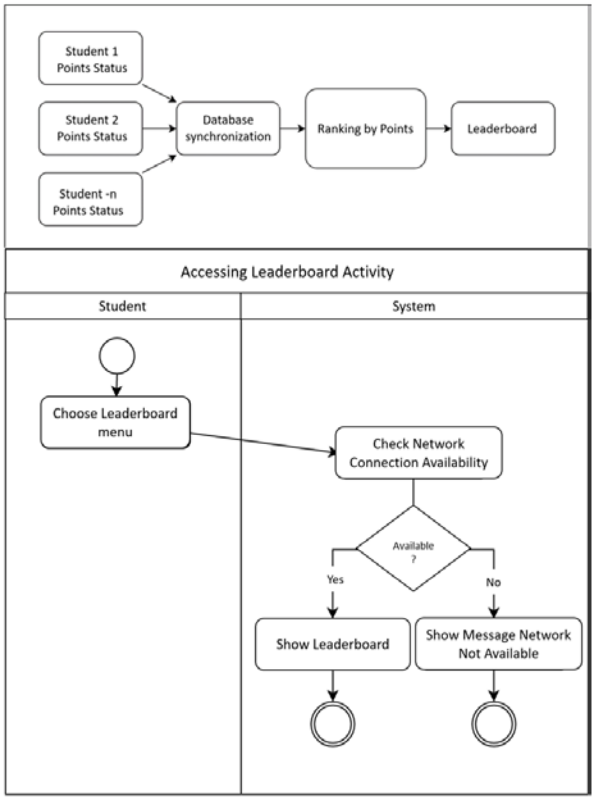


Fig. 23. Leaderboard Flow and Activity Diagram.

6) Avatar flow design: The avatar in the application can change according to the student's level status. Changes in avatars depend on student learning achievements, namely tutorials, evaluations, and points obtained. The following is the avatar flow for the application to be built, which can be seen in Fig. 24.

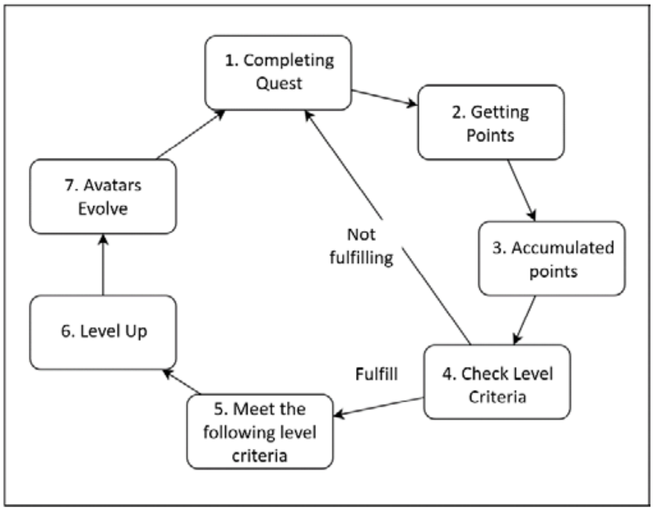


Fig. 24. Avatar Evolution Flow.

Fig. 25 shows student activities to customize their avatar. Avatars can be changed visually according to the level, badges, and rewards that students have.

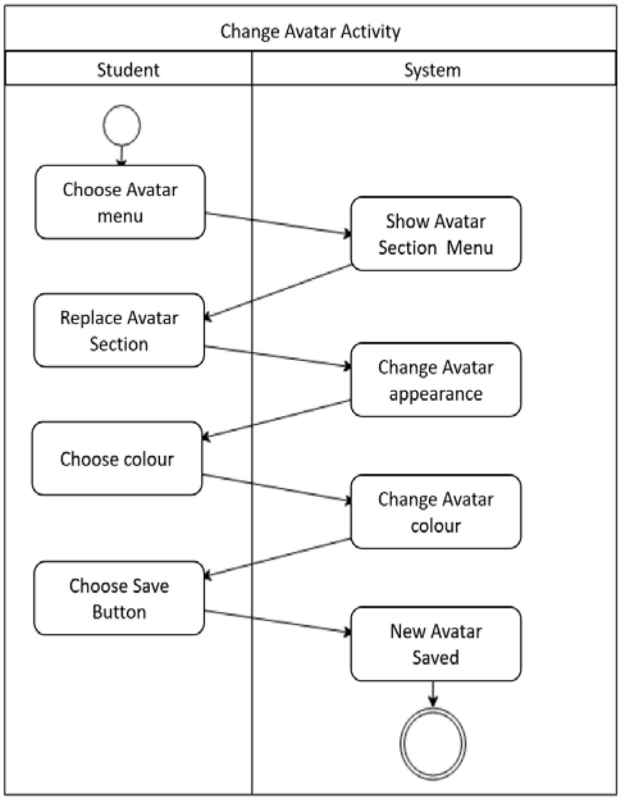
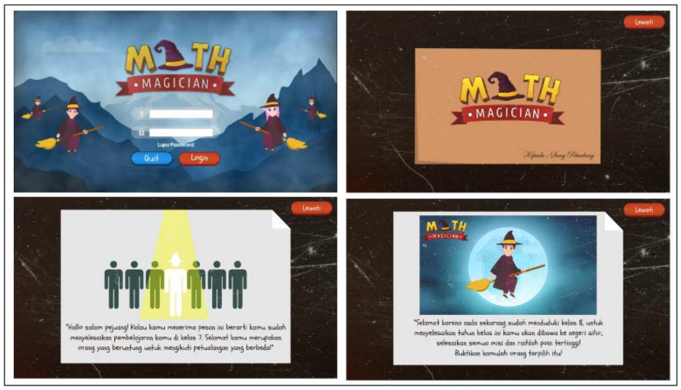


Fig. 25. Avatar Change Activity Diagram.

G. Gamification Asset Design

Assets in the application are symbols that are used to show the concept of gamification in the application that is made. Fig. 26 shows the assets used to deliver students understand the application that will be used.



"Hallo salam pejuang! Kalau kamu menerima pesan ini berarti kamu sudah menyelesaikan pembelajaran kamu di kelas 7. Selamat kamu merupakan orang yang beruntung untuk mengikuti petualangan yang berbeda!"

"Selamat karena anda sekarang sudah menduduki kelas 8, untuk menyelesaikan tahun kelas ini kamu akan dibawa ke negeri sihir, selesaikan semua misi dan raihlah poin tertinggi!
Buktikan kamulah orang terpilih itu!

Fig. 26. Introduction Gamification Asset.

Meanwhile, Fig. 27 is the badge assets that are used to show the form of rewards for student achievement in this gamification application.

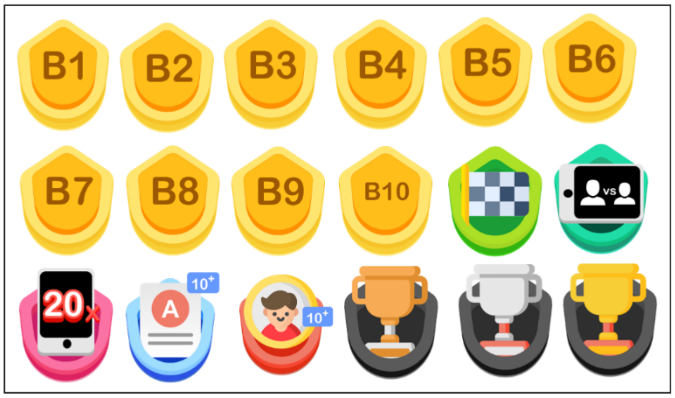


Fig. 27. Badges Gamification Asset.

Fig. 28 shows the assets associated with the Level which is divided into three where each level has a sub-level. The level is adjusted to the achievement of the learning modules that have been completed by students.

Fig. 29 shows the avatar assets that can be used by students after completing the learning stages in the application. The avatar designed in the application consists of shapes that represent the level of the game and can evolve when students advance to the next level.

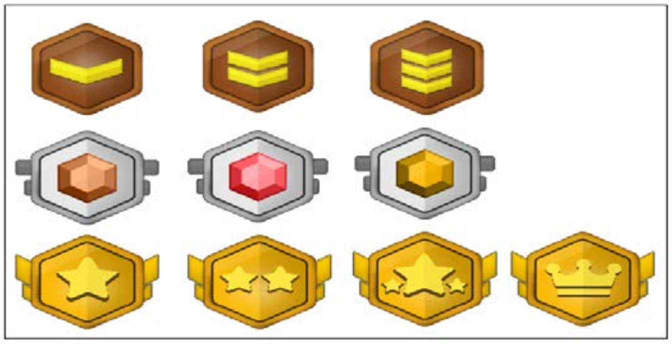


Fig. 28. Level Gamification Asset.

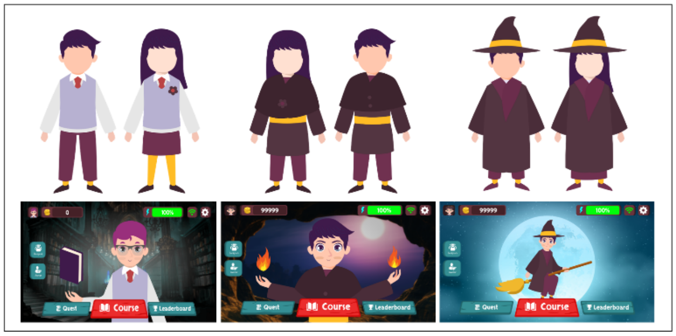
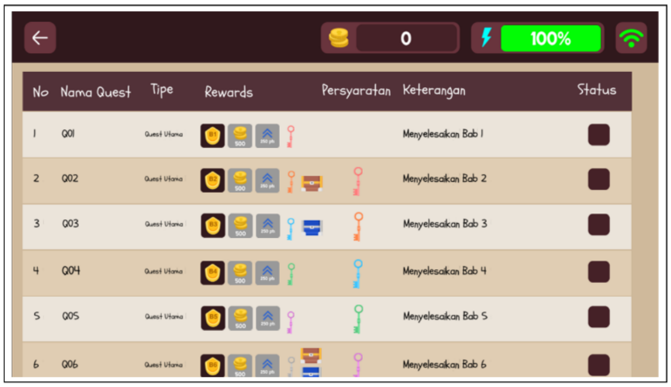


Fig. 29. Avatar Gamification Asset.

Fig. 30 shows quest assets where students can find out the flow of the game, challenges that must be completed, as well as rewards are given along with the status of achievements that have been completed.



No	Nama Quest	Tipe	Rewards	Persyaratan	Keterangan	Status
1	Q01	Quest Utama	B1 500 250 pts		Menyelesaikan Bab 1	
2	Q02	Quest Utama	B2 500 250 pts		Menyelesaikan Bab 2	
3	Q03	Quest Utama	B3 500 250 pts		Menyelesaikan Bab 3	
4	Q04	Quest Utama	B4 500 250 pts		Menyelesaikan Bab 4	
5	Q05	Quest Utama	B5 500 250 pts		Menyelesaikan Bab 5	
6	Q06	Quest Utama	B6 500 250 pts		Menyelesaikan Bab 6	

Fig. 30. Quest Gamification Asset.

Fig. 31 shows the learning assets owned by the application. Where there are features of access to materials, exercises, and evaluations. The form of the material is presented in a video tutorial, while for evaluation, it uses text and images.

Fig. 32 shows the practice and exam assets owned by the application, which are used by students to practice both independently and in multiplayer. Each question will be made random so that students will have different practice questions and exams.

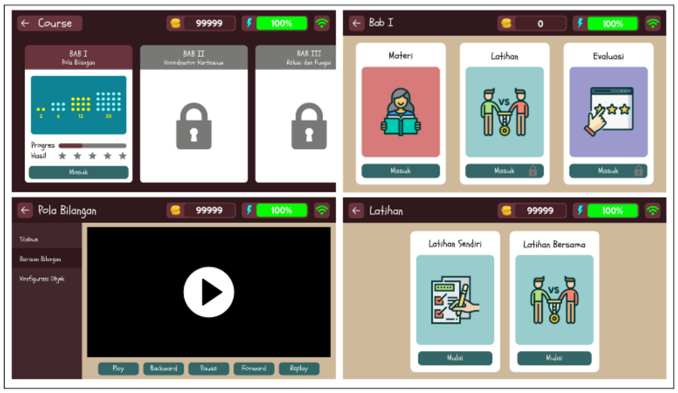


Fig. 31. Learning Materials Gamification Asset.

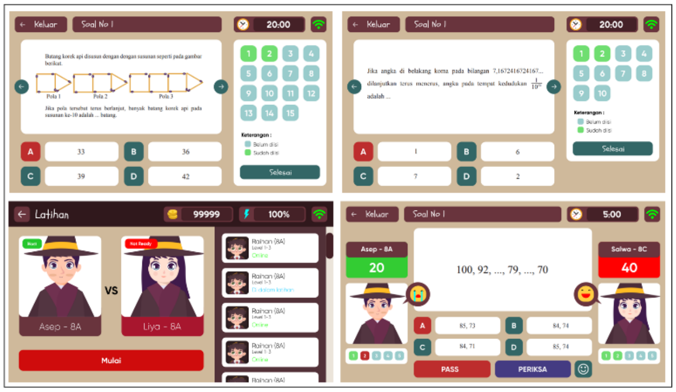


Fig. 32. Practice and Exam Gamification Asset.

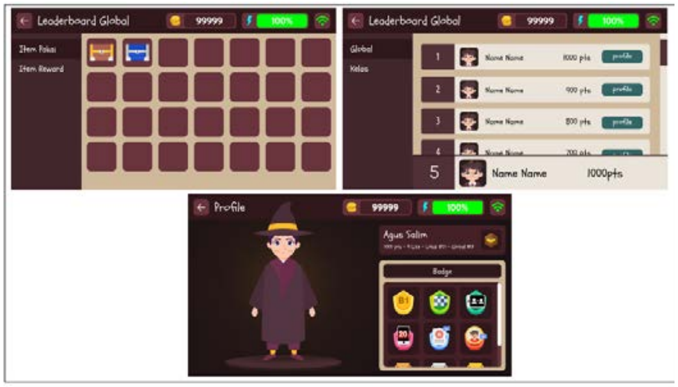


Fig. 33. Profile and Leaderboard Gamification Asset.

Fig. 33 shows the profile assets and leaderboard in the application, this feature is used to see what items have been collected as a form of reward for completing the material and to see the student leaderboard and the points they have.

H. Application Testing

The tests carried out on the application consist of functional (alpha), beta and quasi-experimental testing. The method used in alpha testing is black box testing which focuses on the functional requirements of the system being built and in the beta stage, user assessments of the software are carried out through interviews and questionnaires. Meanwhile, quasi-experimental testing was used to measure the effect of using the application on students' understanding, interest, and motivation.

Alpha testing is a functional test that is used to test the new system. Alpha testing focuses on the functional requirements of the software. Alpha test results can be seen in Table VI.

Beta testing is a test that is carried out objectively, where testing is carried out directly on students. The method used is a questionnaire that is used to conclude the quality assessment of the application being built [49]. The following is the sampling formulation for the questionnaire:

$$n = \frac{N}{1 + Ne^2}$$

n: number of samples

N: Population (421)

e: threshold (10%)

$$n = \frac{421}{1 + 421 * (10\%)^2} = 80,81 \approx 81$$

Based on existing calculations, the number of samples obtained is 81 or equivalent to 2 Grade VIII classes, with an average number of students per class of 42 students. Two test classes were selected, namely grade VIII-A (40 students) and grade VIII-C (41 students) which had 81 students according to the calculation of the number of samples.

The questionnaire was given consists of nine questions related to the quality of the application which can be seen in Table VII, meanwhile for the measurement of results using a Linkert scale with weighted answers which can be seen in Table VIII.

TABLE VI. APPLICATION FUNCTIONAL TESTING

No	Testing components	Testing type	Result
1	Login page	*Black Box*	*Accepted*
2	Narrative Page	*Black Box*	*Accepted*
3	Home page	*Black Box*	*Accepted*
4	Account Settings Page	*Black Box*	*Accepted*
5	Quest Page	*Black Box*	*Accepted*
6	Course Page	*Black Box*	*Accepted*
7	Chapter Page	*Black Box*	*Accepted*
8	Study Page	*Black Box*	*Accepted*
9	Practice Page	*Black Box*	*Accepted*
10	Exam Page	*Black Box*	*Accepted*
11	Backpack Page	*Black Box*	*Accepted*
12	Avatar Creation Page	*Black Box*	*Accepted*
13	Pre-Practice Room Page	*Black Box*	*Accepted*
14	Self-Practice Page	*Black Box*	*Accepted*
15	Joint Practice Page	*Black Box*	*Accepted*
16	Leaderboard Page	*Black Box*	*Accepted*
17	Profile Page	*Black Box*	*Accepted*

TABLE VII. QUESTIONNAIRE QUESTIONS

No	Questions
1	Is the application easy to use?
2	Is the use of colors, buttons, and letters for the appearance (interface) of the application attractive?
3	Does the app encourage me to study math?
4	Can the application help me in understanding math lessons?
5	Is the material presented following the lesson given?
6	Is the material presented interesting and easy to understand?
7	Does using a ranking system motivate me to study?
8	Does the group practice system make learning interesting?
9	Does the use of avatars make the learning system interesting?

TABLE VIII. QUESTIONNAIRE ANSWER SCORE

Answer Category	Score
Strongly agree	5
Agree	4
Neutral	3
Disagree	2
Strongly disagree	1

The results of the final application assessment are the overall calculation of the questionnaire results obtained from student answers, which can be seen in Table IX, with the application quality assessment criteria shown in Table X.

TABLE IX. OVERALL QUESTIONNAIRE RESULTS

Category Answers	Score	Frequency of Answers	Total Score
Strongly agree	5	302	1510
Agree	4	246	984
Neutral	3	181	543
Disagree	2	0	0
Strongly disagree	1	0	0
Total		729	3037

P = (3037/3645) x 100% = 83%

TABLE X. APPLICATION QUALITY ASSESSMENT

No	Percentage Value	Criteria
1	0% - 19%	Very bad
2	20% - 39%	Bad
3	40% - 59%	Neutral
4	60% - 79%	Good
5	80% - 100%	Very good

Based on the results of the overall questionnaire, the application criteria showed 83%, which means that the application was in very good criteria.

Meanwhile, to determine the effect of using the application on students (student engagement), quasi-experimental testing was carried out by conducting pre-test and post-test activities through three activities, namely, understanding the material, filling out motivational questionnaires, and filling out interest questionnaires. The results of the pre-test and post-test will be compared using normalized gain score analysis which is calculated using the following formulation:

Normalized gain(g) = (post-test score– pre-test score) / (maximum score - pre-test score)

The results of the quasi-experimental testing activities obtained the results shown in Table XI and Fig. 34.

Based on the results of the tests carried out, it was concluded that there was an increase in student engagement in learning mathematics after using the gamification mechanism in the m-learning application. The increase was found in student understanding by 42%, student interest in applications by 35%, and student motivation by 33%.

TABLE XI. ASSESSMENT OF STUDENT ENGAGEMENT

Student Engagement	**Score**	***Pre-Test***	***Post-Test***	***Gain***
Student Understanding	Average	58,64	72,47	0,42
Student Interest	Average	58,30	73,07	0,35
Student Motivation	Average	58,84	72,57	0,33

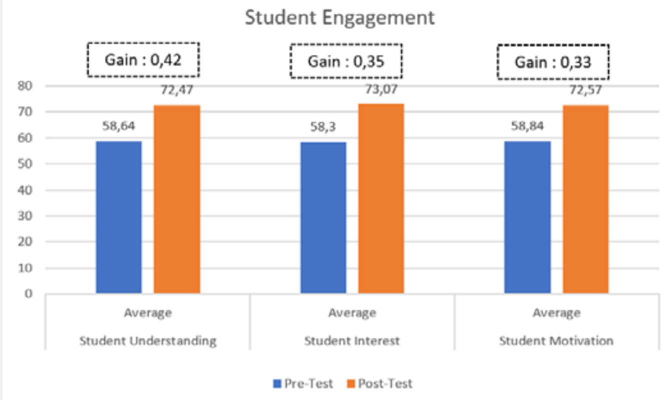


Fig. 34. Results of the Student Engagement Test on the Application.

This study is in line with research [50] where gamification can create student engagement in the learning process, as well as research [51] where there is an increase in better understanding by using gamification-based applications. Gamification-based mobile applications are also an effective and efficient alternative learning media in addition to classroom learning, because the game model can be used anywhere [52], besides that students have a competitive sense in learning because this gamification model shows rankings between participants [53].

The limitations of this research are the absence of features that facilitate schools to update the content in the application if there is a curriculum change, the ability of the application to be able to monitor and provide intelligent direction to students and teachers, and the data storage mechanism is still simple.

Therefore, the development of this research in the future will be involving elements such as artificial intelligence [54][55] to improve students' predictive abilities, in addition to applying augmented reality technology for application interaction [56] as well as having the ability to store and track data and student progress securely using blockchain technology [57][58].

V. CONCLUSION

In this study, we developed a mathematical mobile learning application that combined elements of gamification with the octalysis method and the ARCS learning model.

The purpose of this research is to see how far the role of the application can provide students' engagement in mathematics. The results obtained indicate that there is an increase in understanding by 42%, student interest by 35%, and student motivation by 33% after using the application. This is measured by a pre and post-test mechanism that is carried out on several students who use the application.

As future work, we plan to develop this application by adding interactive modules to adjust the curriculum changes that occur. Addition of intelligent monitoring and evaluation modules can monitor and provide referrals for students and teachers effectively and efficiently. Develop module content with technologies such as augmented reality, as well as secure data storage mechanisms that are easy to trace and transparent using blockchain technology.

ACKNOWLEDGMENT

We would like to thank SMP Negeri 1 Cangkuang, Bandung Regency - West Java, as a place for testing and implementing mathematical gamification applications, as well as the Universitas Komputer Indonesia (UNIKOM) for providing funding for the development of this research.

REFERENCES

[1] K. Pia, “Barriers in teaching learning process of mathematics at secondary level: A quest for quality improvement,” Am. J. Educ. Res., vol. 3, no. 7, pp. 822–831, 2015.

[2] B. R. Acharya, “Factors affecting difficulties in learning mathematics by mathematics learners,” Int. J. Elem. Educ., vol. 6, no. 2, pp. 8–15, 2017.

[3] G. A. Tularam, “Traditional vs Non-traditional Teaching and Learning Strategies-the case of E-learning!,” Int. J. Math. Teach. Learn., vol. 19, no. 1, pp. 129–158, 2018.

[4] K. Das, “İntegrating e-learning & technology in mathematics education,” J. Inf. Comput. Sci., vol. 11, no. 1, pp. 310–319, 2021.

[5] S. Syafril, Z. Asril, E. Engkizar, A. Zafirah, F. A. Agusti, and I. Sugiharta, “Designing prototype model of virtual geometry in mathematics learning using augmented reality,” in Journal of Physics: Conference Series, 2021, vol. 1796, no. 1, p. 12035.

[6] J. Y. Ahn and A. Edwin, “An e-learning model for teaching mathematics on an open source learning platform,” Int. Rev. Res. open Distrib. Learn., vol. 19, no. 5, 2018.

[7] A. Qashou, “Influencing factors in M-learning adoption in higher education,” Educ. Inf. Technol., vol. 26, no. 2, pp. 1755–1785, 2021.

[8] M. Munoto, M. S. Sumbawati, and S. F. M. Sari, “The Use of Mobile Technology in Learning With Online and Offline Systems,” Int. J. Inf. Commun. Technol. Educ., vol. 17, no. 2, pp. 54–67, 2021.

[9] V. Matzavela and E. Alepis, “M-learning in the COVID-19 era: physical vs digital class,” Educ. Inf. Technol., vol. 26, no. 6, pp. 7183–7203, 2021.

[10] P. Yaniawati, I. I. Supianti, D. Fisher, and N. Sa’adah, “Development and effectiveness of mobile learning teaching materials to increase students’ creative thinking skills,” in Journal of Physics: Conference Series, 2021, vol. 1918, no. 4, p. 42081.

[11] X. Yang, X. Zhao, X. Tian, and B. Xing, “Effects of environment and posture on the concentration and achievement of students in mobile learning,” Interact. Learn. Environ., vol. 29, no. 3, pp. 400–413, 2021.

[12] A. N. Saleem, N. M. Noori, and F. Ozdamli, “Gamification Applications in E-learning: A Literature Review,” Technol. Knowl. Learn., no. January, 2021, doi: 10.1007/s10758-020-09487-x.

[13] O. D. Rozhenko, A. D. Darzhaniya, V. V. Bondar, and M. V. Mirzoian, “Gamification of education as an addition to traditional educational technologies at the university,” CEUR Workshop Proc., vol. 2914, pp. 457–464, 2021.

[14] B. Huang and K. F. Hew, “Using Gamification to Design Courses: Lessons Learned in a Three-year Design-based Study,” Educ. Technol. Soc., vol. 24, no. 1, pp. 44–63, 2021.

[15] S. A. A. Freitas, A. R. T. Lacerda, P. M. R. O. Calado, T. S. Lima, and E. D. Canedo, “Gamification in education: A methodology to identify student’s profile,” in 2017 IEEE Frontiers in Education Conference (FIE), 2017, pp. 1–8.

[16] H. A. Yamani, “A Conceptual Framework for Integrating Gamification in eLearning Systems Based on Instructional Design Model,” Int. J.

Emerg. Technol. Learn., vol. 16, no. 4, pp. 14–33, 2021, doi: 10.3991/ijet.v16i04.15693.

[17] A. Bahri and S. Supriyadi, "The Influence of Attention, Relevance, Confidance, and Satisfaction (ARCS) Learning Model on Science Learning Outcomes of Fifth Grade Elementary School Students," Acad. Open, vol. 4, pp. 10–21070, 2021.

[18] L. Ma and C. S. Lee, "Evaluating the effectiveness of blended learning using the ARCS model," J. Comput. Assist. Learn., vol. 37, no. 5, pp. 1397–1408, 2021, doi: 10.1111/jcal.12579.

[19] J. Cabero-Almenara and R. Roig-Vila, "The motivation of technological scenarios in Augmented Reality (AR): Results of different experiments," Appl. Sci., vol. 9, no. 14, 2019, doi: 10.3390/app9142907.

[20] A. Khaleghi, Z. Aghaei, and M. A. Mahdavi, "A gamification framework for cognitive assessment and cognitive training: qualitative study," JMIR serious games, vol. 9, no. 2, p. e21900, 2021.

[21] C. Gellner, I. Buchem, and J. Müller, "Application of the Octalysis Framework to Gamification Designs for the Elderly," in European Conference on Games Based Learning, 2021, pp. 260–XIX.

[22] W. PUSPITARINI, "Customer Motivation Analysis on Retail Business with Octalysis Gamification Framework," J. Theor. Appl. Inf. Technol., vol. 99, no. 13, 2021.

[23] O. Azouz and Y. Lefdaoui, "Gamification design frameworks: a systematic mapping study," in 2018 6th International Conference on Multimedia Computing and Systems (ICMCS), 2018, pp. 1–9.

[24] A. Mora, D. Riera, C. González, and J. Arnedo-Moreno, "Gamification: a systematic review of design frameworks," J. Comput. High. Educ., vol. 29, no. 3, pp. 516–548, 2017.

[25] M. Rauschenberger, A. Willems, M. Ternieden, and J. Thomaschewski, "Towards the use of gamification frameworks in learning environments," J. Interact. Learn. Res., vol. 30, no. 2, pp. 147–165, 2019.

[26] L. da Rocha Seixas, A. S. Gomes, and I. J. de Melo Filho, "Effectiveness of gamification in the engagement of students," Comput. Human Behav., vol. 58, pp. 48–63, 2016.

[27] A. Manzano-León et al., "Between level up and game over: A systematic literature review of gamification in education," Sustainability, vol. 13, no. 4, p. 2247, 2021.

[28] I. Bouchrika, N. Harrati, V. Wanick, and G. Wills, "Exploring the impact of gamification on student engagement and involvement with e-learning systems," Interact. Learn. Environ., vol. 29, no. 8, pp. 1244–1257, 2021.

[29] J. Kim and D. M. Castelli, "Effects of gamification on behavioral change in education: A meta-analysis," Int. J. Environ. Res. Public Health, vol. 18, no. 7, 2021, doi: 10.3390/ijerph18073550.

[30] S. Z. Wahid, "The Effectiveness of Gamification in Improving Student Performance for Programming Lesson The Effectiveness of Gamification in Improving Student Performance for Programming Lesson," no. April, pp. 0–7, 2019.

[31] K. Li and J. M. Keller, "Use of the ARCS model in education: A literature review," Comput. Educ., vol. 122, pp. 54–62, 2018.

[32] R. W. Pratama, S. Sudiyanto, and R. Riyadi, "The Development Of Attention, Relevance, Confidence, And Satisfaction (ARCS) Model Based on Active Learning to Improve Students' learning Motivation," Al-Jabar J. Pendidik. Mat., vol. 10, no. 1, pp. 59–66, 2019.

[33] A. M. Afjar and M. Syukri, "Attention, relevance, confidence, satisfaction (ARCS) model on students' motivation and learning outcomes in learning physics," in Journal of Physics: Conference Series, 2020, vol. 1460, no. 1, p. 12119.

[34] A. S. A. A. Sharma, "Integrating the ARCS Model with Instruction for Enhanced Learning," J. Eng. Educ. Transform., vol. 32, no. 3, 2019.

[35] Y. Kim, N. S. Yu, and G. Lee, "Development of Teaching-Learning Plans Applying ARCS Motivation Strategies for Food Safety Education," J. Korean Home Econ. Educ. Assoc., vol. 31, no. 3, pp. 135–153, 2019.

[36] Y. Chou, Actionable gamification: Beyond points, badges, and leaderboards. Packt Publishing Ltd, 2019.

[37] A. J. Irawan, F. A. T. Tobing, and E. E. Surbakti, "Implementation of Gamification Octalysis Method at Design and Build a React Native Framework Learning Application," in 2021 6th International Conference on New Media Studies (CONMEDIA), 2021, pp. 118–123.

[38] V. Yfantis and D. Tseles, "Exploring Gamification In The Public Sector Through The Octalysis Conceptual Model," 2017.

[39] A. C. Luther, Authoring interactive multimedia. Academic Press Professional, Inc., 1994.

[40] A. H. Sutopo, "Multimedia interaktif dengan flash," Yogyakarta Graha Ilmu, pp. 32–48, 2003.

[41] T. Busono, N. D. Herman, E. Krisnanto, J. Maknun, and N. I. K. Dewi, "Luther's model implementation on multimedia development for building construction subject in vocational high school (SMK)," in 5th UPI International Conference on Technical and Vocational Education and Training (ICTVET 2018), 2019, pp. 334–338.

[42] T. Wibowo and V. Limken, "Designing Learning Media For Bataknese Cuisine Using Multimedia Development Life Cycle (Mdlc) Method," J. Inf. Syst. Technol., vol. 2, no. 2, pp. 56–63, 2021.

[43] S. Purwanti, R. Astuti, J. Jaja, and R. Rakhmayudhi, "Application of the Multimedia Development Life Cycle (MDLC) Methodology to Build a Multimedia-Based Learning System," Budapest Int. Res. Critics Inst. Humanit. Soc. Sci., vol. 5, no. 1, pp. 2498–2506, 2022.

[44] H. Putri, I. Shadiq, and G. G. Putri, "Interactive Learning Media for Cellular Communication Systems using the Multimedia Development Life Cycle Model," J. Online Inform., vol. 6, no. 1, pp. 1–10, 2021.

[45] I. P. Sari, A. Jannah, A. Syahputra, and R. Tanjung, "Vector Analysis of the Prayer Movement on Health Using Visual Media Multimedia Application Development Life Cycle," Indones. J. Educ. Soc. Sci. Res., vol. 2, no. 1, pp. 147–157, 2021.

[46] S. L. Rahayu and R. Dewi, "Educational Games as A learning media of Character Education by Using Multimedia Development Life Cycle (MDLC)," in 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018, pp. 1–4.

[47] A. Heryandi and I. Afrianto, "Online Diploma Supplement Information System Modelling for Indonesian Higher Education Institution," in IOP Conference Series: Materials Science and Engineering, 2019, vol. 662, no. 2, doi: 10.1088/1757-899X/662/2/022092.

[48] I. Afrianto, A. Heryandi, A. Finandhita, and S. Atin, "Prototype of E-Document Application Based on Digital Signatures to Support Digital Document Authentication," IOP Conf. Ser. Mater. Sci. Eng., vol. 879, no. 1, 2020, doi: 10.1088/1757-899X/879/1/012042.

[49] H. L. Tuan, C. C. Chin, and S. H. Shieh, "The development of a questionnaire to measure students' motivation towards science learning," Int. J. Sci. Educ., vol. 27, no. 6, pp. 639–654, 2005, doi: 10.1080/0950069042000323737.

[50] I. Bouchrika, N. Harrati, V. Wanick, and G. Wills, "Exploring the impact of gamification on student engagement and involvement with e-learning systems," Interact. Learn. Environ., vol. 29, no. 8, pp. 1244–1257, 2021, doi: 10.1080/10494820.2019.1623267.

[51] M. A. C. Madrid and D. M. A. de Jesus, "Towards the Design and Development of an Adaptive Gamified Task Management Web Application to Increase Student Engagement in Online Learning," in International Conference on Human-Computer Interaction, 2021, pp. 215–223.

[52] W. Ortiz, D. Castillo, and L. Wong, "Mobile Application: A Serious Game Based in Gamification for Learning Mathematics in High School Students," in 2022 31st Conference of Open Innovations Association (FRUCT), 2022, pp. 220–228.

[53] N. P. Harvey Arce and A. M. Cuadros Valdivia, "Adapting competitiveness and gamification to a digital platform for foreign language," Int. J. Emerg. Technol. Learn., vol. 15, no. 20, pp. pp. 194–209, Oct. 2020.

[54] K. Duggal, L. R. Gupta, and P. Singh, "Gamification and machine learning inspired approach for classroom engagement and learning," Math. Probl. Eng., vol. 2021, 2021.

[55] Y. R. Pratama, S. Atin, and I. Afrianto, "Predicting Student Interests against Laptop Specifications through Application of Data Mining Using C4.5 Algorithms," in IOP Conference Series: Materials Science and Engineering, 2019, vol. 662, no. 2, doi: 10.1088/1757-899X/662/2/022129.

[56] I. Afrianto, A. F. Faris, and S. Atin, "Hijaiyah letter interactive learning for mild mental retardation children using Gillingham method and augmented reality," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 6, pp. 334–341, 2019, doi: 10.14569/ijacsa.2019.0100643.

[57] I. Afrianto, T. Djatna, Y. Arkeman, I. Hermadi, and I. S. Sitanggang, "Block Chain Technology Architecture For Supply Chain Traceability Of Fisheries Products In Indonesia: Future Challenge," J. Eng. Sci. Technol. Spec. Issue INCITEST2020, pp. 41–49, 2020.

[58] I. Afrianto and Y. Heryanto, "Design and Implementation of Work Training Certificate Verification Based On Public Blockchain Platform," in 2020 Fifth International Conference on Informatics and Computing (ICIC), 2020, pp. 1–8.

Energy-based Collaborative Filtering Recommendation

Tu Cam Thi Tran[1, [0000-0001-5811-6952]], Lan Phuong Phan[2], Hiep Xuan Huynh[3*, [0000-0002-9213-131X]]

Faculty of Information Technology, Vinh Long University of Technology Education (VLUTE), Vinh Long province, Vietnam[1]
College of Information and Communication Technology, Can Tho University (CTU), Can Tho city, Vietnam[2, 3]

***Abstract*—The core value of the recommendation model is the using of the measures to measure the difference between the jumps (e.g. pearson), some other studies based on the magnitude of the angle in space (e.g. cosine), or some other studies study the level of confusion (e.g. entropy) between users and users, between items and items. Recommendation model provides an important feature of suggesting the suitable items to user in common operations. However, the classical recommendation models are only concerned with linear problems, currently there is no research about nonlinear problems on the basis of potential/energy approach to apply for the recommendation model. In this work, we mainly focus on applying the energy distance measure according to the potential difference with the recommendation model to create a separate path for the recommendation problem. The theoretical properties of the energy distance and the incompatibility matrix are presented in this article. Two experiment scenarios are conducted on Jester5k, and Movielens datasets. The experiment result shows the feasibility of the energy distance measures/ the potential in the recommendation systems.**

Keywords—Energy distance; energy model; collaborative filtering; recommendation system; distance correlation; incompatibility

I. Introduction

Recommendation systems suggest the suitable items to a user based on his/her purchased items or his/her rated items [9]. There are many implementations of a recommendation system based on different factors and applied to different contexts, such as the recommendation systems determining the user's rating values according to the magnitude of the angle (e.g. cosine) [22], or the recommendation systems based on the difference of the users (e.g. pearson) [19][2], or the recommendation systems based on the confuse of one user with another (e.g. entropy) [4], some other recommendation systems based on the statistical implication [12][18].

Collaborative filtering recommendation model [6] mainly based on users, items. In particular, the Singular Value Decomposition algorithm - a classical method from linear algebra used as a technique to reduce size in machine learning - is combined with recommendation model, or Alternating Least Squares (ALS) - a matrix factorization algorithm – is used for the larger-scale collaborative filtering problems, or some techniques for selecting random or popular items are also integrated to recommendation systems. However, most of the recommendation models revolve around the problems of linear relations, not the problems of nonlinear relations.

In this article, we propose a new collaborative filtering recommendation model to consider nonlinear relations instead of focusing only on linear relations between users. This approach is performed on the basis of determining the relationship/distance between users in pairs, especially Newton's gravitational potential energy (known as potential energy, shortly energy) between two users. In this collaborative filtering recommendation model, the relationship between two users is determined through calculating the maximum mean discrepancy (MMD), or a lack of compatibility or similarity between two or more users.

The article is structured as follows. In Section II, we present collaborative filtering based on energy. Section III presents the learning model, data division methods and evaluation methods. In Section IV, we propose the new recommendation model based on energy. In Section V, we show the experiment on the Jester5k, and MovieLense datasets. Section VI is the conclusion of the article.

II. Collaborative Filtering

Collaborative filtering [3][15][22] is the process of filtering or evaluating items using the opinions of others. Collaborative filtering technology gathers the opinions of large interconnected communities on the webs, and supports filtering of substantial quantities of data. The recommendation system [1] uses a lot of information such as: the items, the users and the rating values to suggest the suitable items to user. However, the unwanted information has been removed by using the computerized methods before presenting the recommendation result to the user.

In collaborative filtering [1], the recommendation system searches for similar users to make predictions. The user's rating model is a useful feature for determining similarity. Normally, the collaborative filtering recommendation methods use ratings without additional information about the user or the item to recommend the suitable items.

The recommendation system [9][10][11] is an information system [17], includes a set of four:

$$S = < U, I, R, f > \tag{1}$$

Where

U - is the set of users (the closed universe), $U = \{u_1, u_2, \ldots, u_n\}$ with $u_k \in U, k = 1..n$. U is a finite set of n objects (a nonempty set)

*Corresponding Author.

I - is the set of items $I = \{i_1, i_2, \ldots, i_m\}$ with $i_j \in I, j = 1..m$. I is a finite set of m attributes (a nonempty set).

$R = \{r_{ij}\}$, with $i = 1..n$, $j = 1..m$. R is a rating matrix, where r_{ij} is rating value of the user u_i to item i_j.

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & .. & r_{1m} \\ r_{21} & r_{22} & r_{23} & .. & r_{2m} \\ r_{31} & r_{32} & r_{33} & .. & r_{3m} \\ .. & .. & .. & .. & .. \\ r_{n1} & r_{n2} & r_{n3} & .. & r_{nm} \end{bmatrix}$$

For example, Table I is a rating matrix where the rating value ranges from 1 to 5 or not available.

$f : U \, x \, I \rightarrow$ R - is the total decision function called the information function such that $f(u, i) \in R_i$ for every $i \in I$, $u \in U$. Function $f(u_k, i_j)$ is used to measure the relevance (the rating value) of item i_j with user u_k. The rating value $f : U \, x \, I \rightarrow R$.

A. Energy Distance

The energy distance [7][8][13] is the distance between the probability distributions. Energy is defined as the similarity in the form of potential energy between objects in gravitational space. The potential energy is zero if and only if the positions (centers of gravity) of the two objects coincide, and the potential energy increases as the difference between the objects in space increases. The concept of potential energy can be applied to collaborative filtering. Let U_1 and U_2 be independent random vectors in U , where F and G are cumulative distribution functions, and they correspond to each other. Accordingly, $\|.\|$ represents the Euclidean normal of its argument, E represents the expected value, and a random variable U_1' represents a copy (iid), which is independent and distributed like U_1; that mean, U_1 and U_1' are iid. Similarly, U_2 and U_2' are iid. The squared energy distance [16][20][24] can be determined according to the expected distance between random vectors.

$$D^2(F, G) := 2E\|U_1 - U_2\| - E\|U_1 - U_1'\| - E\,\|U_2 - U_2'\| \geq 0 \qquad (2)$$

Consider the null hypothesis that two random variables, U_1 and U_2, have the same cumulative distribution functions: F = G. For samples $u_{11}, \ldots, u_{1n}$ from U_1 and $u_{21}, \ldots, u_{2m}$ from U_2, respectively, the E-statistic for testing this null hypothesis is:

$$\varepsilon_{n,m}(U_1, U_2) := 2A - B - C \qquad (3)$$

where A, B, and C are simply averages of pairwise distances:

$$A = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \|u_{1i} - u_{2j}\|, \qquad (4)$$

$$B = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|u_{1i} - u_{1j}\|, \qquad (5)$$

$$C = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \|u_{2i} - u_{2j}\|: \qquad (6)$$

One can prove (3) that $\varepsilon(U_1, U_2) := D^2(F, G)$ is zero if and only if U_1 and U_2 have the same distribution $(F = G)$. It is also true that the statistic $\varepsilon_{n,m}$ is always non-negative. When the null hypothesis of equal distributions is true, the test statistic.

$$T = \frac{nm}{n+m} \varepsilon_{n,m}(U_1, U_2) \qquad (7)$$

B. Incompatibility Matrix

The incompatibility matrix E represents the energy distance between users. The incompatibility matrix $E = \{e_{ij}\}$, with $i = 1..n$, $j = 1..n$, e_{ij} is calculated by formula (2).

$$E = \begin{bmatrix} e_{11} & e_{12} & e_{13} & .. & e_{1n} \\ e_{21} & e_{22} & e_{23} & .. & e_{2n} \\ e_{31} & e_{32} & e_{33} & .. & e_{3n} \\ .. & .. & .. & .. & .. \\ e_{n1} & e_{n2} & e_{n3} & .. & e_{nn} \end{bmatrix}$$

For example, Table II shows the matrix representing the energy distance between users by using information of Table I.

C. Incompatibility Neighborhood

The neighborhood of the user u_a is defined by the energy distance between the users and u_a . The neighborhood is filtered with a certain number of the users, who has the lowest potential energy (i.e. k nearest neighbors - knn).

TABLE I. AN EXAMPLE OF THE RATING MATRIX R

	$\mathbf{i_1}$	$\mathbf{i_2}$	$\mathbf{i_3}$	$\mathbf{i_4}$	$\mathbf{i_5}$	$\mathbf{i_6}$	$\mathbf{i_7}$	$\mathbf{i_8}$
u_1	?	4.0	4.0	2.0	1.0	2.0	?	?
u_2	3.0	?	?	?	5.0	1.0	?	?
u_3	3.0	?	?	3.0	2.0	2.0	?	3.0
u_4	4.0	?	?	2.0	1.0	1.0	2.0	4.0
u_5	1.0	1.0	?	?	?	?	?	1.0
u_6	?	1.0	?	?	1.0	1.0	?	1.0
u_7	1	3.0	2.0	?	?	2.0	?	?
u_8	5	?	?	2.0	1.0	?	?	?
u_9	?	4.0	?	?	1.0	2.0	?	?

TABLE II. AN EXAMPLE OF THE MATRIX OF ENERGY DISTANCE FOR THE ACTIVE USERS

	$\mathbf{u_1}$	$\mathbf{u_2}$	$\mathbf{u_3}$	$\mathbf{u_4}$	$\mathbf{u_5}$	$\mathbf{u_6}$	$\mathbf{u_7}$	$\mathbf{u_8}$	$\mathbf{u_9}$
u_1	0	0	0	0	0	0	0	0	0
u_2	7.874	0	0	0	0	0	0	0	0
u_3	7.280	5.196	0	0	0	0	0	0	0
u_4	8.306	6.403	2.828	0	0	0	0	0	0
u_5	6.000	5.656	4.795	5.385	0	0	0	0	0
u_6	5.567	5.196	4.898	5.830	1.732	0	0	0	0
u_7	3.316	6.557	6.324	6.928	3.605	3.464	0	0	0
u_8	7.810	5.000	4.000	4.690	4.795	5.656	6.164	0	0
u_9	4.472	6.480	6.708	7.549	4.000	3.316	2.645	7.000	0

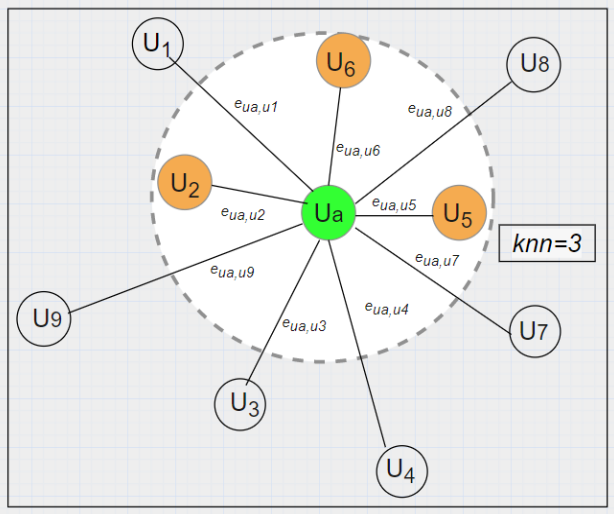


Fig. 1. The Neighborhood of u_a with knn = 3.

To find the k-nearest neighbors (knn) for u_a, the energy distance is used. Fig. 1 shows the 2D space of the incompatibility points with the active user u_a - the users with low energy will be displayed closer together. If knn equals to 3, u_2, u_5 and u_6 are selected to be three nearest neighbors of u_a.

D. Rating prediction

The predicted rating $\hat{r}_{aj}$ of user $u_a$for item i_j is calculated by (8),

$$\hat{r}_{aj} = \frac{1}{\sum_{i \in N(a)} e_{ai}} \sum_{i \in N(a)} e_{ai}\, r_{ij} \tag{8}$$

where: e_{ai} is the incompatibility between u_a and the user u_i in the neighborhood. $N(a)$ is knn of the user u_a. r_{ij} is the rating value of the user u_i to item i_j.

E. Top N Items Recommendation

To recommend the suitable items to the active user u, N items with the highest ranking are selected. The ranking of each item i is calculated by the ranking function. This function is reversible to map the predicted rating on the normalized scale back to the original rating scale. Normalization is used to remove individual rating bias by users who use lower or higher ratings than other users. A popular method is to center the rows of the rating matrix by formula:

$$h(r_{ui}) = \hat{r}_{ui} - \bar{r}_u \tag{9}$$

Where $\bar{r}_u$ is the average of all available ratings in row u (i.e. the available ratings of user u) of the rating matrix R; $\hat{r}_{ui}$ is the predicted rating of user u to item i.

III. Recommendation Evaluation

A. K-folds Cross Evaluation

In order to evaluate the effectiveness of recommender models [5][25], k-folds cross evaluation method is performed. In this article, the dataset is divided into a training set and a testing set with k-folds = 5. The dataset is splitted into 5 subsets, all subsets of equal size, 80% (4 subsets) of the dataset is used for training and 20% (1 subset) of the dataset is used for testing. The model is evaluated recursively 5 times, each time using a different train/test split, which ensures that all users and items are considered for both training and testing. The results are then averaged to produce the final result.

B. Evaluation

To evaluate the recommendation model, three measures of error: 1. Mean Absolute Error (MAE); 2. Mean Squared Error (MSE); and 3. Root Mean Square Error (RMSE) is used. The evaluation of the error of the recommendation model is an important step in the design of the recommendation system. This helps the designer to select the model and they can check the error of the model before the designer applies this model in practice.

- Root Mean Square Error (RMSE) [14][23]. Root mean square error between real rating value r_{ij} and the predicted rating value $\hat{r}_{ij}$ is calculated by fomular (10).

$$RMSE = \sqrt{\frac{\sum_{(i,j) \in n} (r_{ij} - \hat{r}_{ij})^2}{|n|}} \tag{10}$$

- Mean Squared Error (MSE) [14][23]. The mean square error between real rating value r_{ij} and the predicted rating value $\hat{r}_{ij}$ is calculated by fomular (11).

$$MSE = \frac{\sum_{(i,j) \in n} (r_{ij} - \hat{r}_{ij})^2}{|n|} \tag{11}$$

- Mean Absolute Error (MAE) [14][23]. The mean absolute error between real rating value r_{ij} and the predicted rating value $\hat{r}_{ij}$ is calculated by fomular (12).

$$MAE = \frac{1}{|n|} \sum_{(i,j) \in n} \left| r_{ij} - \hat{r}_{ij} \right| \tag{12}$$

IV. Energy based Recommendation Model

A. Model

Fig. 2 presents the general overview of energy-based recommendation model including four components: the dataset (U x I x R), energy including incompatibility matrix which is calculated by the energy distance measure, predicted ratings performing the prediction, and the table of predicted rating to be used for recommending top n items to active user u_a.

The dataset consists of a set of m items (items), where $I = \{i_1, i_2, \dots, i_j, i_{j+1}, \dots, i_m\}$ and the set n users (users), where $U = \{u_1, u_2, u_k, u_{k+1} \dots, u_n\}$, the rating values $R = \{r_{11}, r_{12}, \dots, r_{1n};\ r_{21}, r_{22}, \dots, r_{2n}; \dots; r_{m1}, r_{m2}, \dots, r_{mn}\}$, u_a is the active user to be recommended.

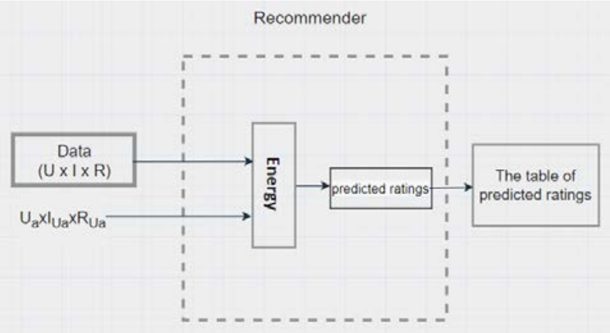


Fig. 2. Energy-based Recommendation Model.

B. Algorithm

The recommendation algorithm of the energy based model includes six steps as the folow:

```
Algorithm. Energy-based recommendation

    Input: The data matrix (U x I x R); and the active user needs to be suggested $u_a$
    Output: The rating prediction table to be used for recommendation to the active users $u_a$;
Begin
Step 1: Calculating the incompatibility matrix E by using the energy distance of $u_a$ with all users
Step 2: Finding k nearest neighbors (int u, int i, int [][]R, int [][]E)
    // u is the users, i is the items, rating is the rating matrix, E is the incompatibility matrix.
    if (R[u][i] != 0 && E [$u_a$][ u] != 0)
Step 3: Predicting the rating value of $u_a$ for items based on k nearest neighbors.
        <$R^{'}$[a][j] = (1/E($u_a$, u))(E($u_a$, u) $R^{'}$[i][j])>
Step 4: Calculating the ranking of each item <List_N[i]>;
Step 5: Sorting the list of predicted ratings in descending order < Sort (List_N)>;
Step 6: Recommeding the top $N$ item with the highest ranking to the active $u_a$ < Print (Top-N>);
End.
```

V. Experiment

A. Datasets

Experiment is performed on the Jester5k and MovieLense datasets. These two datasets are summarized in the Table III, and the distribution of ratings of them is displayed Fig. 3.

Jester5k[1] contains the ratings of 5000 anonymous users collected from the Jester Online Joke Recommendation System between April 1999 and May 2003. This data set contains 5000 users and 100 jokes with ratings ranging from -10.00 to +10.00. All selected users have rated 36 or more jokes.

MovieLense [2] (100k) were published in 1998 by GroupLense (https://grouplens.org). This dataset includes 100,000 (100k) ratings from 943 users for 1682 movies with ratings ranging from 1 to 5. Each user has rated at least 20 movies.

B. Tool

The "recommenderlab" package [21] is used in experiment of this article; specifically, user based collaborative filtering model using cosine measure (named UBCFCosine RS).

TABLE III. THE TABLE TO DESCRIBE DATASETS: JESTER 5K AND MOVIELENSE

Names	Number of rows (users)	Number of col.s (items)	Number of ratings	Value domain of ratings
Jester5k	5000	100	362106	-10 - +10
MovieLense	943	1682	99392	1 - 5

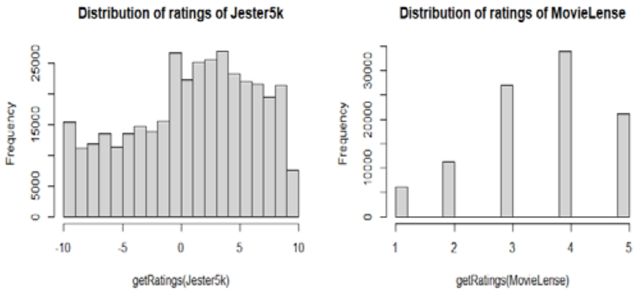


Fig. 3. Distribution of Ratings of Jester5k vad Movielense Datasets.

We have implemented the proposed energy based recommendation model (named UBCFEnergy RS) in R language. This model is integrated into the recommenderlab package. We have also written a function to compare the results of the proposed model UBCFEnergy RS and the selected model of recommenderlab package UBCFCosine RS.

C. Scenario 1: Recommendation on Jesterk5k

This scenario evaluates the errors (MAE, RMSE, MSE) of two recommendation models UBCFEnergy RS and UBCFCosine RS.

The comparison results of errors (MAE, MSE, RMSE) of the two models are shown in Fig. 4 for each known ratings (given = 2, 16, 36) on all k nearest neighbors knn = 10, 20, 30, 40. The results show that the error of the UBCFEnergy RS model is always smaller than that of UBCFCosine RS model.

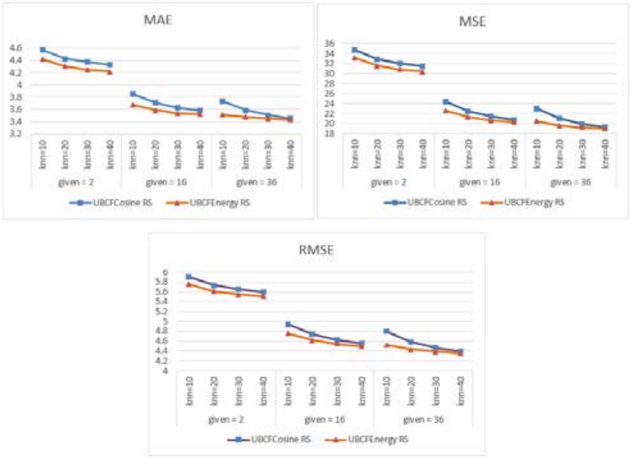


Fig. 4. Errors for each given (2, 16, 36) on all knn = 10, 20, 30, 40.

[1] https://rdrr.io/cran/recommenderlab/man/Jester5k.html, accessed on February 01, 2021.

[2] https://rdrr.io/cran/recommenderlab/man/MovieLense.html

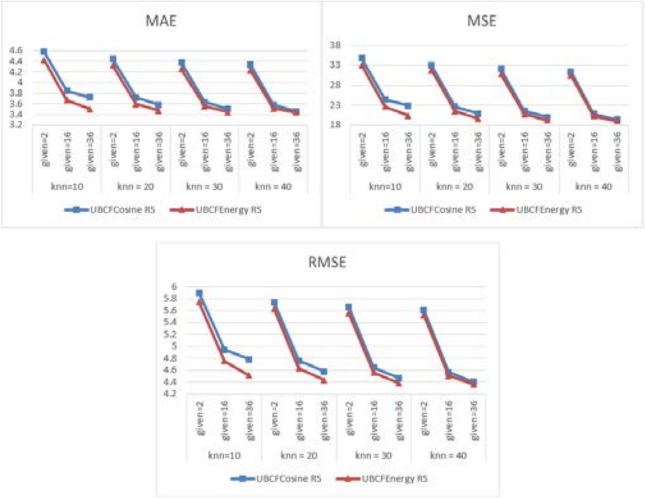


Fig. 5. Errors with each knn (10, 20, 30, 40) for all given = 2, 16, 36.

Fig. 5 presents the errors of UBCFEnergy RS and UBCFCosine RS for each k nearest neighbors knn = 10, 20, 30, 40 on all known ratings (given = 2, 16, 36). The experiment result also show that the proposed model is better than UBCFCosine RS model.

D. Scenario 2: Recommendation on MovieLense

This scenario presents the experement result of two models UBCFEnergy RS and UBCFCosine RS on MovieLense dataset.

Fig. 6 shows the comparison results of errors (MAE, MSE, RMSE) of the two models for each known ratings (given = 4, 10, 20) on all k nearest neighbors knn = 10, 20, 30, 40. Fig. 7 presents the errors of two models for each k nearest neighbors knn = 10, 20, 30, 40 on all known ratings (given = 4, 10, 20). Both results indicates that the errors of UBCFEnergy RS model are smaller than those of UBCFCosine RS model.

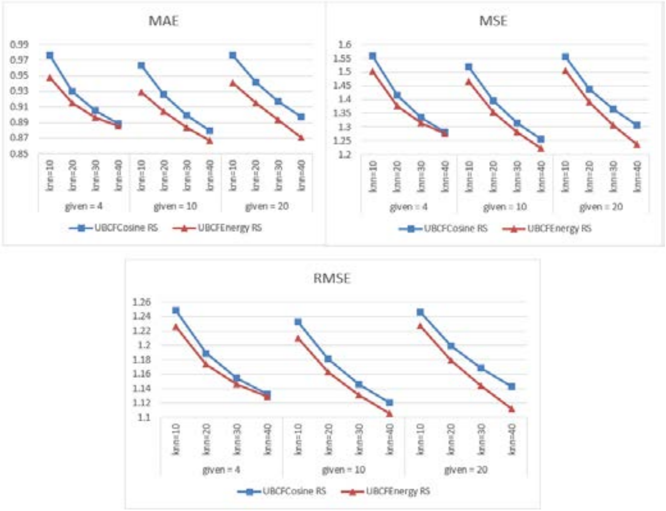


Fig. 6. Errors for each given (4, 10, 20) on all knn = 10, 20, 30, 40.

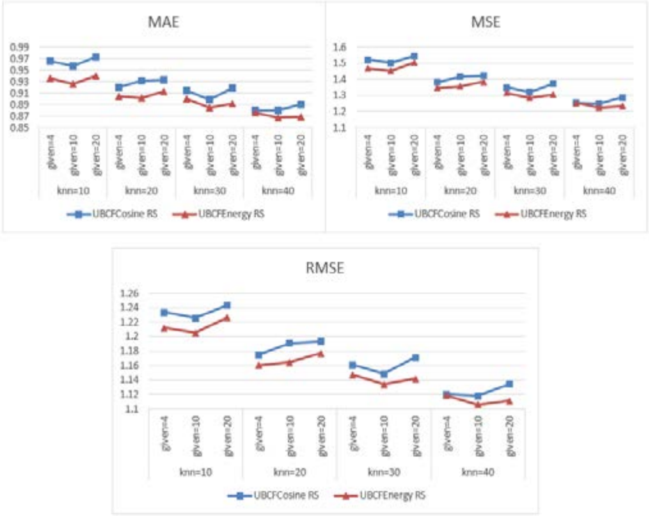


Fig. 7. Errors with each knn (10, 20, 30, 40) for all given = 4, 10, 20.

VI. Conclusion

We have built a new recommendation model based on energy UBCFEnergy RS. The errors (MAE, MSE, RMSE) of this model are compared with the error of user based collaborative filtering model using cosine measure of recommenderlab" package UBCFCosine RS on Jester5k and MovieLense datasets, two datasets commonly used in evaluating the effectiveness of recommendation models. The experimental results of the proposed recommendation model have the lower errors than the compared model on both Jester5k and Movielense. Therefore, the energy-based recommendation model shows the feasibility of applying the energy distance to build the recommendation systems.

References

[1] A.B. Suhaim, and J. Berri, “Context-Aware Recommender Systems for Social Networks: Review, Challenges and Opportunities,” in IEEE Access, vol 9, pp. 57440-57463, 2021.

[2] ABMK. Hossain, Z. Tasnim, S. Hoque, S. Hoque, and M.A. Rahman, “A Recommender System for Adaptive Examination Preparation using Pearson Correlation Collaborative Filtering,” Int J Auto AI Mach Learn, vol 21, pp. 30-43, 2021.

[3] B. Hong, and M. Yu, “A collaborative filtering algorithm based on correlation coefficient. Neural Computing and Applications, vol 31, pp. 8317–8326, 2019.

[4] C. Hemalatha, and B. Bharat, “Personalized recommender system using entropy based collaborative filtering technique,” Journal of Electronic Commerce Research, vol 12, 2011.

[5] C.C. Aggarwal, Recommender Systems, vol. 1. Springer, Heidelberg 2016.

[6] D. Jannach, P. Pu, F. Ricci, and M. Zanker, “Recommender Systems: Past, Present, Future,” AI Magazine, vol 42, 2021. pp. 3–6.

[7] D. Edelmann, T F. Móri, and G. J. Székely, On relationships between the Pearson and the distance correlation coefficients, Statistics & Probability Letters, Vol 169, 108960, ISSN 0167-7152, (2021).

[8] E. Martínez-Gómez, M.T. Richards, and DStP. Richards, “Distance correlation method for discovering associations in large astrophysical databases,” the Astrophysical Journal, American Astronomical Society, vol 781, 2014.

[9] G. Adomavicius, and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” IEEE Transactions on Knowledge and Data Engineering, vol. 17, pp. 734–749, 2005.

[10] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin, “Incorporating contextual information in recommender systems using a multidimensional approach,” ACM Trans. Inf. Syst, Vol 23, pp. 103–145, 2005.

[11] G. Adomavicius, N. Manouselis, and Y. Kwon, Multi-criteria recommender systems, Recommender Systems Handbook, pp.769-803 2011.

[12] H.X. Huynh, Q.N. Phan, T.N. Duong, and T.T.H. Nguyen, “Collaborative Filtering Recommendation Based on Statistical Implicative Analysis,” International Conference on Computational Collective Intelligence, Springer, Cham, pp. 224-235, 2020.

[13] H. Zhang, Y. Jian, and P. Zhou, “Collaborative Filtering Recommendation Algorithm Based on Class Correlation Distance,” Recent Advances in Computer Science and Communications, vol 14, pp. 887–894, 2021.

[14] J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J.T. Riedl, “Evaluating collaborative filtering recommender systems,” ACM Trans. Inf. Syst, vol 22, pp. 5–53, 2004.

[15] J. Chen, C. Zhao, Uliji, and L. Chen, “Collaborative filtering recommendation algorithm based on user correlation and evolutionary clustering. Complex & Intelligent Systems. 6(1), pp. 147–156 (2020).

[16] J.S. Gábor, L.R. Maria, and K.B. Nail, “Measuring and testing dependence by correlation of distances,” The Annals of Statistics. Institute of Mathematical Statistics, vol 35, pp. 2769-2794, 2007.

[17] K.J. Cios, W. Pedrycz, and RW. Swiniarski, “Data mining: knowledge discovery methods,” Hardback, Book. English. Published New York; London: Springer, pp. 29-37, 2007.

[18] L.P Phan, H.H. Huynh, and H.X. Huynh, “Hybrid recommendation based on implicative rating measures,” Proceedings of the 2nd International Conference on Machine Learning and Soft Computing, (ICMLSC 2018), ACM, pp. 50-56, 2018.

[19] L. Sheugh, and S.H. Alizadeh, “A note on pearson correlation coefficient as a metric of similarity in recommender system,” 2015 AI & Robotics (IRANOPEN), pp. 1-6, 2015.

[20] M. Rizzo, and G. Székely, “Energy distance. Wiley Interdisciplinary Reviews,” Computational Statistics, vol 8, pp. 27-38, 2016.

[21] M. Hahsler, recommenderlab: “A Framework for Developing and Testing Recommendation Algorithm”, 2015.

[22] S. Ramni, M. Sargam, T. Tanisha, N. Tushar, and S. Gaurav, “Movie Recommendation System using Cosine Similarity and KNN,” International Journal of Engineering and Advanced Technology, Vol 9, pp. 2249-8958, 2020.

[23] S. Ben, J. Ben, F. Dan, Dan, Herlocker, Jon, Shilad, and S. Shilad, “Collaborative Filtering Recommender Systems,” The Adaptive Web, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, vol 4321, pp. 291-32, 2007.

[24] T. Park, X. Shao, and S. Yao, “Partial martingale difference correlation,” Electronic Journal of Statistics, vol 9, pp. 1492-1517, 2015.

[25] Y. Koren, and R. Bell, Advances in Collaborative Filtering, Recommender Systems Handbook, Springer, pp. 145-186, 2011.

Grape Leaves Diseases Classification using Ensemble Learning and Transfer Learning

Andrew Nader, Mohamed H.Khafagy, Shereen A. Hussien
Department of Computer Science
Faculty of Computers and Information
Fayoum University
Egypt

***Abstract*—Agriculture remains an important sector of the economy. Plant diseases and pests have a big impact on plant yield and quality. So, prevention and early detection of crop disease are some of the measures that must be implemented in farming to save the plants at an early stage and thereby reduce the overall food loss. Grapes are the most profitable fruit, but they are also vulnerable to a variety of diseases. Black Measles, Black Rot, and Leaf Blight are diseases that affect grape plants. Manual disease diagnosis can result in improper identification and use of pesticides, and it takes a long time. A variety of deep learning approaches have been used to address this issue of the identification and classification of grape leaf diseases. However, there are also limits to such approaches. Therefore, this paper uses deep learning with the concept of ensemble learning based on three famous Convolutional Neural Network (CNN) architectures (Visual Geometry Group (VGG16), VGG19, and Extreme Inception (Xception)). These three models are pre-trained with ImageNet. The performance of the proposed approach is analyzed using the Plant Village (PV) dataset of common grape leaf diseases. The Proposed model gives higher performance than the results achieved by using each Deep Learning architecture separately and compared with the recent approaches in this study. The proposed system outperformed the others with 99.82% accuracy.**

***Keywords*—Ensemble learning; grape leaf diseases; convolutional neural network (CNN); transfer learning**

I. Introduction

Agriculture is one of the most important sectors in the world, providing humans with food, raw materials, and other essentials. According to the United Nations (FAO) Food and Agriculture Organization, the world's population will reach 9.1 billion by 2050. As a result, to meet the nutrient needs of such a large population, the food growth rate needs to be boosted to 70% by 2050 [1]. Plant disease is one of the leading causes of crop loss. A plant disease is an abnormal condition that modifies the appearance or function of a plant. The visible effects of the disease on plants are described as symptoms. A symptom is any detectable change in the plant's color, form, and/or functions in response to a pathogen or disease-causing agent. The leaf is the most essential portion of the plant to inspect for illness. Plant leaf diseases are divided into three categories: fungal, viral, and bacterial [12, 13].

A. Fungal Diseases

Parasitic organisms have a system of branching threads that make up their bodies. Fungi can enter the host via stomata, but many can pierce a solid surface. Once within, they can either expand through the live cells or stay primarily in the gaps between them. Black Rot, Black Measles, Leaf Blight, and Leaf Rust are signs of these diseases.

B. Viral Diseases

Viruses are not dispersed by water or wind, unlike bacteria and fungi. But insects and worms are the primary vectors of viral pathogens to plants.

C. Bacterial Diseases

Bacteria are small. There are two hundred kinds of bacteria that cause diseases in plants. Their form determines their classification. Spherical, rod-shaped microorganisms and twisted rods are the three primary kinds that may be known. There are many symptoms of a bacterial infection. Leaf Spot is considered the most prevalent of them.

Plant diseases, and how to quickly diagnose and address them to improve the health of crops, are among the most important problems that face agriculture and have an impact on its trade. Identifying the disease before it spreads over the farm to other plants and treating it is a massive challenge in and of itself. It takes a lot of time and effort to determine what kind of disease it has. Furthermore, not all disease types can be accurately identified by the farmer's naked eye.

Initially, plant infections can be detected with the help of an agricultural professional who is familiar with plant diseases. However, manual plant disease identification and determination is a strenuous task and takes a long time. A person's knowledge and experience determine the accuracy of a manual prediction [2].

To overcome the above problems, many studies on machine learning (ML)-based technologies, such as support vector machines (SVM) [3], K-nearest neighbors (KNN) [4], have recently been applied to improve decision-making to classify plant diseases [5]. However, all these proposed approaches face several challenges, including identifying regions of concern for processing and analysis (feature representation). Now, Deep Learning (DL) is considered the next evolution of machine learning, such as convolutional neural networks (CNN) [6]. This work proposes an approach for classifying grape plant diseases based on Ensemble Learning that aggregates three customized CNN architectures with trained weights (VGG16, VGG19, and Xception). The following are the study's major contributions:

- Enhance classification performance.
- Reduce overfitting.

The rest of this paper is structured as follows: In Section II, related works present the most common plant disease terminologies and techniques. The proposed model and materials are illustrated in Section III. Section IV discusses the results of the proposed model and compares them with other related models. Section V, the conclusion. Section VI, the future work.

II. Related Work

Many methods for disease classification from the planet leaf have been presented, especially on the Plant village dataset. The Plant village Dataset [7] is a well-known, publicly available crop leaf collection that contains thousands of images. It has been used in various research and achieved good results. Some of these are as follows:

Akshai KP et al. [8] proposed a method to classify images of grape plant diseases from the Plantvillage dataset using the trained model. The CNN, VGG19, ResNet-152v2, and DenseNet models are all trained. The DenseNet model was the most accurate, with a score of 98.27%. For the convolution layer, a rectified linear activation function, or ReLU activation function, is utilized, and for the output layer, a Softmax activation function is used. The images were reduced to 224x224 pixels using Kera's image data generator, then augmentations like rotation, zoom, and shift were added. At a ratio of 80:20, the dataset is split into training and validation sets. This work had the problem of splitting the dataset into training and validation only and testing the model with the same data, which led to overfitting.

Y. Nagaraju et al. [9] proposed a fine-tuned VGG-16 network to categorize eight different apple and grape leaf types together. The Kera’s library is used to load the pre-trained VGG-16 network. In a typical model, the SoftMax (classifier) layer is removed, and a new output layer (classifier) with a SoftMax activation function is added. The disease dataset for apples and grape leaves is split 80:20 with an accuracy of 97.87%. This work had the problem that although using 30 epochs, the accuracy was 97.87%.

E. Hirani et al. [10] proposed deep learning methods for identifying plant diseases. A Plantvillage dataset is used at a ratio of 80:20, with 70295 images for training and 17572 images for validation. This work used three methods: a customized convolutional neural network, INCEPTIONv3, Small Transformer Network (STN), and Large Transformer Network (LTN), with an accuracy of 95.566%, 97.14%, 97.66%, and 97.98%. This work had a problem with the resolution of images at 256*256 that needed more computation and time in the training and testing phases.

K. Z. Thet et al. [11] proposed a fine-tuning VGG16 with the GAP layer instead of VGG16’s two fully connected layers before the SoftMax layer to classify the diseases on grape leaves. This work has achieved 98.4% more accuracy than others. It mostly focused on five diseases that are prevalent in Myanmar Grapevine Yard. This work had a problem with the overall performance where the number of epochs and batch size were not determined.

III. Proposed Methodology

A. Dataset Description

The Plant Village dataset [7] is used to examine the proposed model performance in the classification of grape leaves diseases. It is a large and freely accessible database. It contains over 55,000 RGB images divided into 38 classes representing 14 different plant species. There are 12 healthy leaf classes as in Fig. 3 and 26 unhealthy leaf classes as in Fig. 1, 2, 4 out of 38 total. The grape leaves images are used in this study, which includes 4,062 images divided into four classes shown in Table I. The following is a summary of each selected disease [13, 14]:

1) Grape: black rot: Black rot fungus attacks on leaves grown in warm, wet seasons, the black rot fungus attacks the upper surface of the leaves, turning them reddish-brown and causing round to angular dots to emerge. As the spots merge, irregular reddish-brown blotches appear as shown in Fig.1.

2) Grape: leaf blight: It is also a fungal disease on leaves grown in high humidity conditions, caused by Exserohilumturcicum. At first, small yellow dots emerge along the leaf margins, then grow to become brown patches as shown in Fig. 4.

3) Grape: esca (black measles): The esca fungus can affect leaves at any time during the growing season, but it is most common in July and August. The symptom is an interveinal "striping". In red varieties, the "stripes" are dark red, while in white cultivars, they are yellow. Fig. 2.

B. Proposed System

In this work, a powerful approach for classification grape leaves diseases based on Ensemble Learning has been presented as shown in Fig. 5.

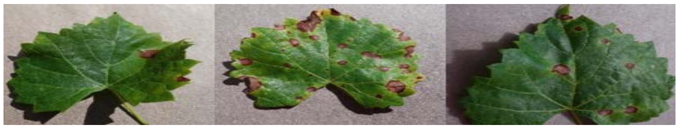


Fig. 1. Grape_Black_Rot.

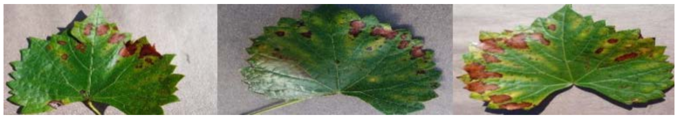


Fig. 2. Grape_ Esca (Black Measles).

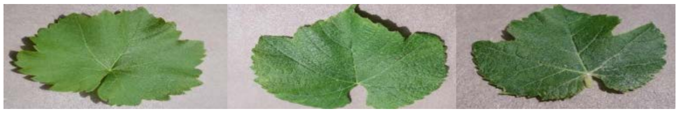


Fig. 3. Grape_Healthy.

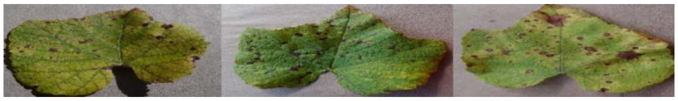


Fig. 4. Grape_ Leaf Blight.

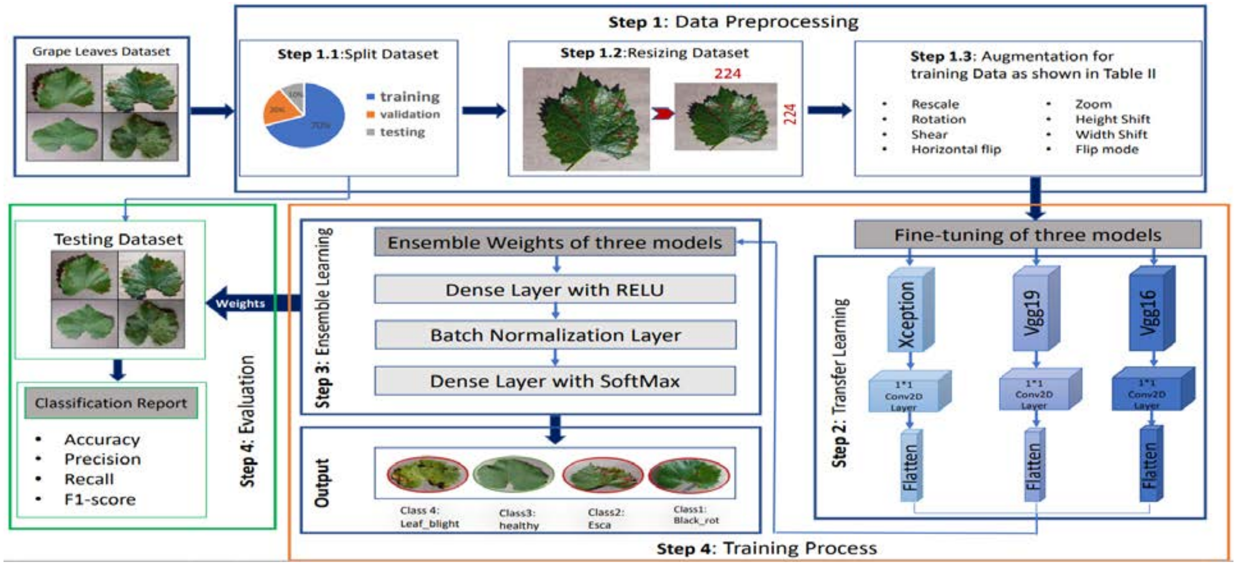


Fig. 5. The Framework of the Proposed Model.

1) Data preprocessing: Preprocessing of the dataset is one of the major roles in training a model. First, divided dataset into three categories: training 70%, validation 20%, and 10% for testing, as shown in Table I. Second, grape images are resized to 224*244*3 pixels. Third, applied data augmentation techniques to generate more training data and avoid overfitting. These techniques include rescale, rotation, shear, horizontal flip, zoom, height shift, width shift, and flip mode on the training set. The data augmentation is applied only to the training data shown in Table II.

TABLE I. DESCRIPTION OF PLANT VILLAGE DATASET

Grape Dataset						
Fig	**Class Name**	**Cause of disease**	**Total samples**	**Training Samples (70%)**	**Validation Samples (20%)**	**Test Samples (10%)**
1	Black Rot	Fungus	1180	826	236	118
2	Esca		1383	969	276	138
3	Leaf blight		1076	754	215	107
4	Healthy	--------	423	296	85	42

TABLE II. AUGMENTATION TECHNIQUES SUMMARY

Techniques	**Values**
Rescale	1./255
Rotation	30
Shear	0.1
Horizontal flip	True
Zoom	0.3
Height shift	0.2
Width shift	0.2
Flip mode	Nearest

2) Model building: Transfer learning is a great approach because it allows to use a pre-trained CNN model with multiple datasets to train a specific dataset.

VGG-16 and VGG-19 [15] refer to the "Visual Geometry Group". They're two different versions of the same structure. The differences between them are as follows, respectively: The VGG-16 consists of 16 layers of the deep neural network, whereas the VGG-19 consists of 19 layers. Both networks contain blocks, where each block is composed of 2D convolution and pooling layers. The Conv2D layer [16] extracts a feature of the image using filters or kernels. The filter is passed throughout the width and height of the input and the dot products function between the input and filter is calculated at every position. Convolution Layer Formula (1).

$$n_{out}= [\frac{n_{in}+2p-k}{s}]+1 \tag{1}$$

It also contains ReLU (Rectified Linear Activation Function) [17, 18] that returns all negative values set to zero (2). The function and gradient in ReLU (2) and (3).

$$\text{ReLU}(x) = \max(0, x) \tag{2}$$

$$\frac{d}{dx}\ ReLU(x)\text{=1 if x > 0; otherwise} \tag{3}$$

The Conv2D layer is followed by a pooling layer to reduce the computation and the number of parameters. Max pooling (4) is one of the most used pooling operations. Then the matrix is flattened into a vector. The flattened vector is passed into the FC (Fully Connected) layer. FC is used to connect each node in one layer to each node in another layer. The last layer is SoftMax. It is located at the end of the FC layer, which predicts a multi-class.

$$\text{S}(x) = \underset{i\,=\,1}{\overset{N}{max}}\ x_i \tag{4}$$

Xception [19] refers to Extreme Inception. First, the data passes through the entering flow, then eight times through the middle flow, and finally through the exit flow. Batch normalization is applied to all convolution and separable convolution layers. Separable convolutions are time-saving and more efficient than classical convolutions.

The proposed model depends on a customized CNN with trained weights from the VGG16, VGG19, and Xception models. The learning scenario starts with receiving grape leaves images from the input layer. The input layer is shared with three pre-trained networks. Three models are reshaped by freezing all their layers. But we removed the top layer (output layer) from each model to add the proposed output layers.

Added two layers, a Conv 1x1 layer with 1024 filters with padding-zero and stride-one, to collect the most important features and allow for reduced dimensionality followed by a flattening layer to convert the matrix to the tensor of one dimension (vector). As illustrated in Fig. 6-8.

3) Ensemble learning: After applying the flattening layer to each model, added the merged layer to aggregate the flattening layer from each network and use it as an input (new input) for the ensemble learning model. Then, added a dense layer that allowed every neuron in this layer to connect to the next layer by weight followed by a batch normalization layer, which allows each layer to make learning more independent. Finally, added the output layer with the SoftMax activation function to predict the final output.

4) Training phase: The proposed model compiled with the Adam optimizer [20] is a stochastic gradient descent method with a learning rate equal 2e-5. The loss function is a categorical cross-entropy [21] that is used in the multi-class classification task equation (5) to calculate the loss. The proposed model is fitted at epochs where value equals 10, train batch size equals 16, and validate batch size equals 8.

$$\text{Loss}=-\sum_{i=1}^{output\ size} y_i . \log \hat{y}_i \tag{5}$$

5) Testing phase: The best-saved weights were loaded after those testing images were loaded. Finally, images were resized to 224*224 and fed the testing images to the model to classify four classes of the grape leaves' diseases.

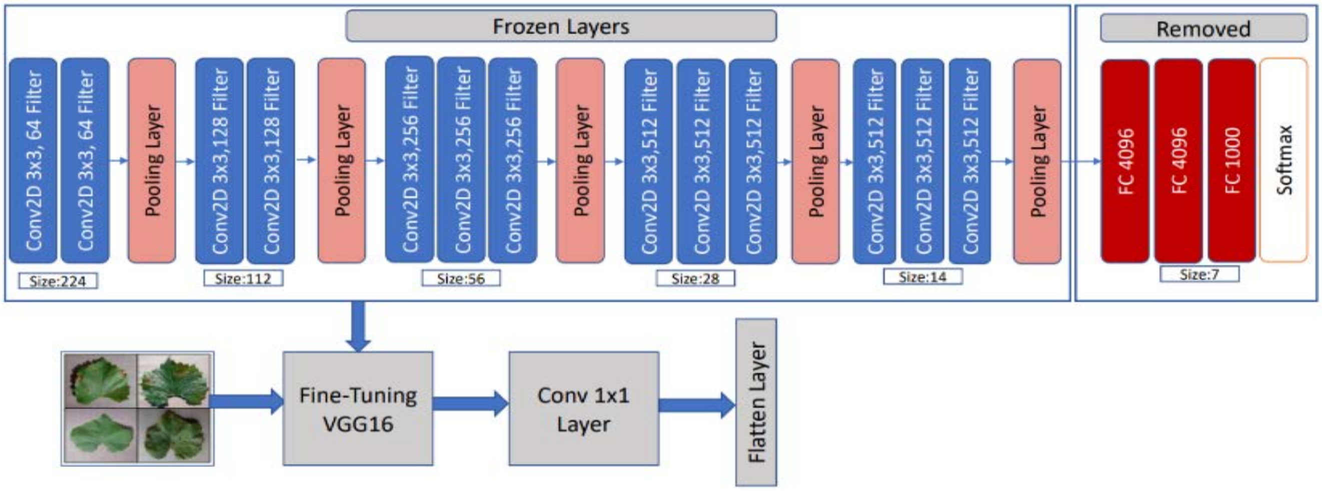


Fig. 6. Fined Tuned for VGG16.

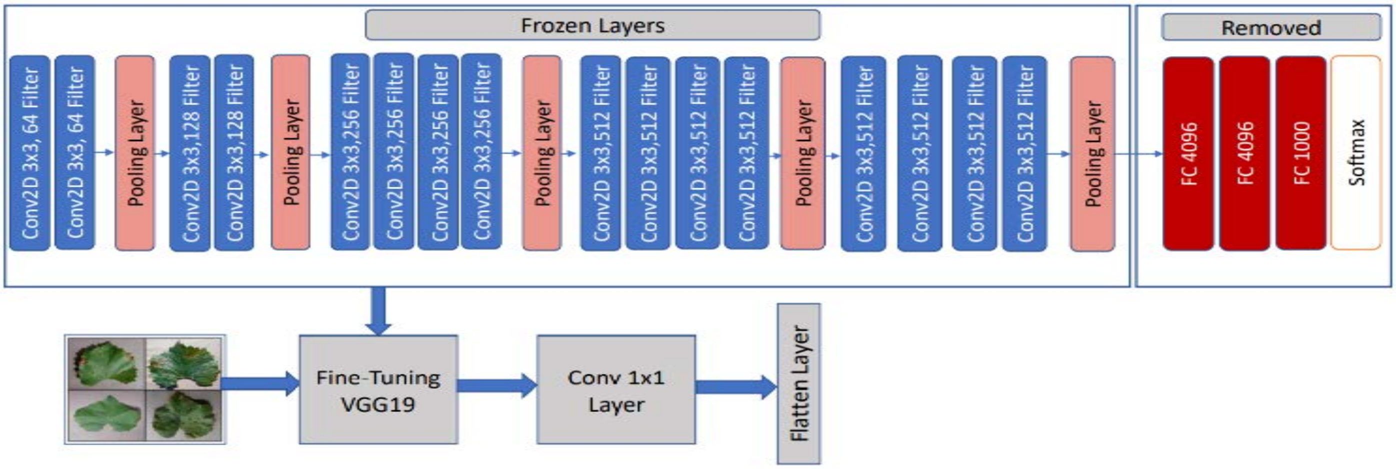


Fig. 7. Fined Tuned for VGG19.

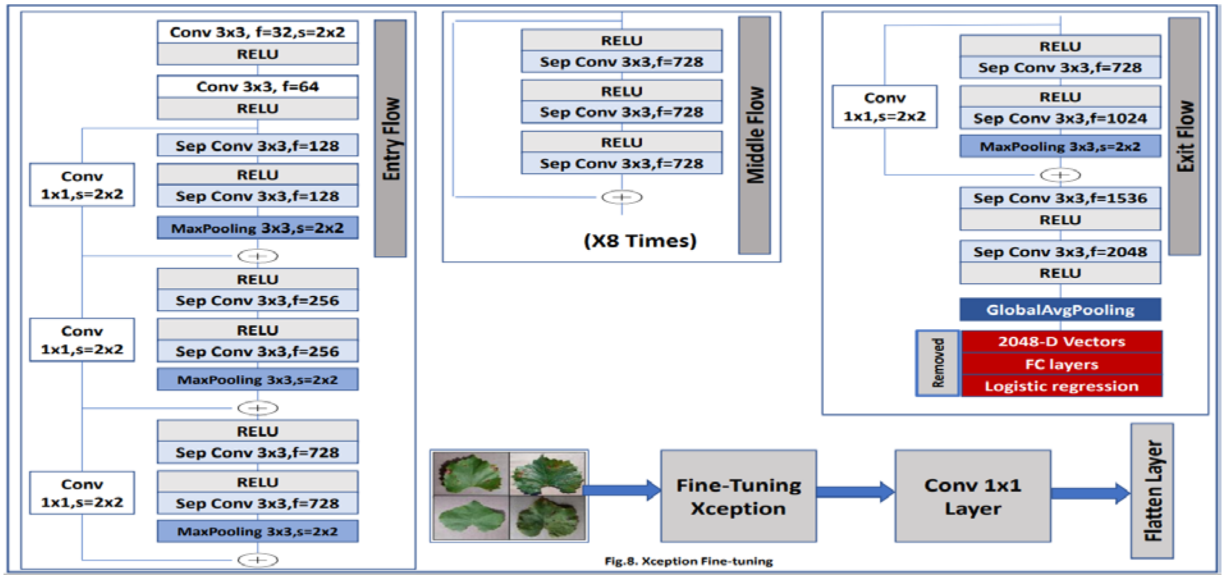


Fig. 8. Fined Tuned for Xception.

IV. Experimental Work & Results

A. Experimental Settings

This work demonstrates that the highest accuracy is achieved with a relatively small dataset of grape leaves in the plant village dataset using ensemble learning (e.g., VGG16 Model, VGG19, and Xception). The dataset is divided into a ratio of 7:2:1 as shown in Table I. All experiments are implemented using Colab [22] provided by Google, the Keras framework [23] that can run on top of TensorFlow, and the Python programming language. All experiments were conducted on a 12 GB NVIDIA Tesla K80 GPU (Graphical Processing Unit) and 12 GB of RAM.

B. Experimental Evaluation

The performance of the proposed model is evaluated using accuracy (9), precision (6), recall (7), F1-score (8), and confusion matrix in the testing phase equations shown in Table III. Moreover, loss, accuracy, validation loss, and validation accuracy are calculated during different epochs in the training phase, as shown in Fig. 10-17. Also, we compared our proposed model performance as shown in Table V with other models (e.g. [8]) that work on the same dataset as shown in Table IV and with each architecture used in the proposed model separately. Summary of the training model in the Fig. 9 to understand the underlying parameters.

TABLE III. Performance Equations Summary

Assessments	Equation	Equ.No
Precision (P)	$\frac{TP}{TP+FP}$	(6)
Recall (R)	$\frac{TP}{TP+FN}$	(7)
F1-Score (F)	$2 * \frac{P*R}{P+R}$	(8)
Accuracy (Acc)	$\frac{TP+TN}{TP+TN+FP+FN}$	(9)

Where:

True Positive (TP) the model predicts the positive class correctly. True Negative (TN) model correctly classifies the negative class. In a false positive (FP), the model predicts the positive class incorrectly. In false negative (FN), the model predicts the negative class incorrectly.

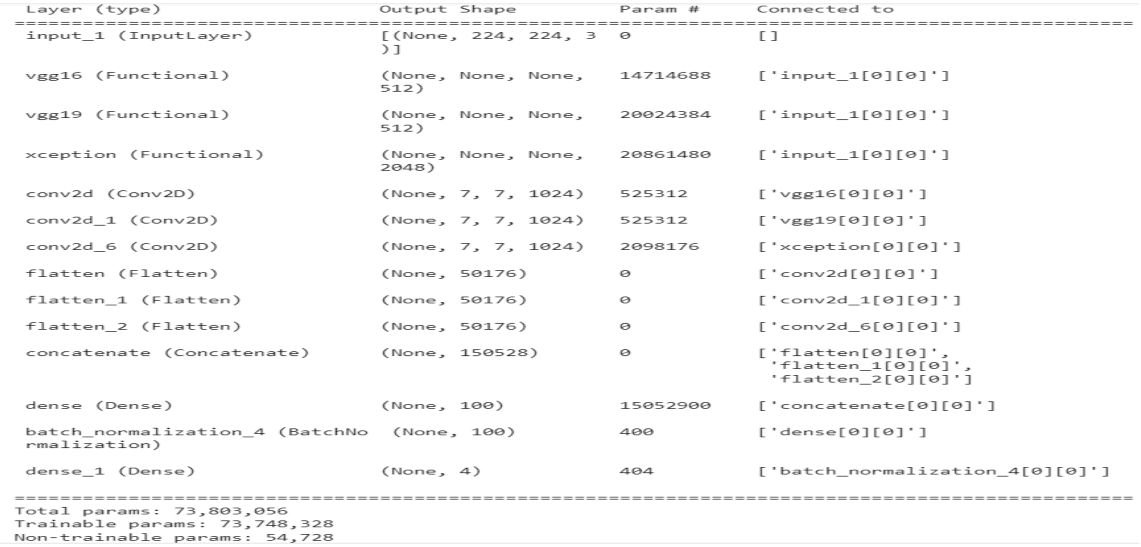
The total training time for the proposed model is about 24 min 28 sec ± 2 min with 10 epochs, and the test time is about 5 seconds. Other models in [8] use 20 epochs in training. However, our proposed method achieved the highest accuracy, showing the feasibility of our proposed method. Compared with existing methodologies, we can see that our proposed model has some degree of competition in terms of accuracy and precision, as shown in Tables IV and V. The robustness of our suggested approach is confirmed by these results.

TABLE IV. Result of Existing Models

Model	Precision	Recall	F1-score	Accuracy	epochs
CNN [8]	94.60	94.58	94.56	94.58	20
VGG [8]	95.54	95.32	95.32	95.32	20
RESNET [8]	97.11	97.04	97.05	97.04	20
DENSENET [8]	98.31	28.27	98.28	98.27	20

TABLE V. CLASSIFICATION REPORT FOR PROPOSED MODELS

Model	Classes	Precision	Recall	F1-score	Support	Accuracy
Fined Tuned VGG16	Black Rot	1.00	0.92	0.96	118	0.96
	Esca	1.00	0.94	0.97	138	
	Healthy	0.98	1.00	0.99	107	
	Leaf blight	0.74	1.00	0.85	42	
Fined Tuned VGG19	Black Rot	1.00	097	0.98	118	0.99
	Esca	0.97	1.00	0.99	138	
	Healthy	1.00	1.00	1.00	107	
	Leaf blight	1.00	1.00	1.00	42	
Fined Tuned Xception	Black Rot	0.98	0.99	0.99	118	0.99
	Esca	0.99	0.99	0.99	138	
	Healthy	1.00	0.99	1.00	107	
	Leaf blight	0.98	1.00	0.99	42	
Ensemble Model (Proposed)	Black Rot	1.00	1.00	1.00	118	1.00
	Esca	1.00	1.00	1.00	138	
	Healthy	1.00	1.00	1.00	107	
	Leaf blight	1.00	1.00	1.00	42	



```
 Layer (type)                    Output Shape          Param #     Connected to
==================================================================================================
 input_1 (InputLayer)            [(None, 224, 224, 3   0           []
                                 )]

 vgg16 (Functional)              (None, None, None,    14714688    ['input_1[0][0]']
                                 512)

 vgg19 (Functional)              (None, None, None,    20024384    ['input_1[0][0]']
                                 512)

 xception (Functional)           (None, None, None,    20861480    ['input_1[0][0]']
                                 2048)

 conv2d (Conv2D)                 (None, 7, 7, 1024)    525312      ['vgg16[0][0]']

 conv2d_1 (Conv2D)               (None, 7, 7, 1024)    525312      ['vgg19[0][0]']

 conv2d_6 (Conv2D)               (None, 7, 7, 1024)    2098176     ['xception[0][0]']

 flatten (Flatten)               (None, 50176)         0           ['conv2d[0][0]']

 flatten_1 (Flatten)             (None, 50176)         0           ['conv2d_1[0][0]']

 flatten_2 (Flatten)             (None, 50176)         0           ['conv2d_6[0][0]']

 concatenate (Concatenate)       (None, 150528)        0           ['flatten[0][0]',
                                                                    'flatten_1[0][0]',
                                                                    'flatten_2[0][0]']

 dense (Dense)                   (None, 100)           15052900    ['concatenate[0][0]']

 batch_normalization_4 (BatchNo   (None, 100)          400         ['dense[0][0]']
 rmalization)

 dense_1 (Dense)                 (None, 4)             404         ['batch_normalization_4[0][0]']

==================================================================================================
Total params: 73,803,056
Trainable params: 73,748,328
Non-trainable params: 54,728
```

Fig. 9. Summary for the Proposed Model.

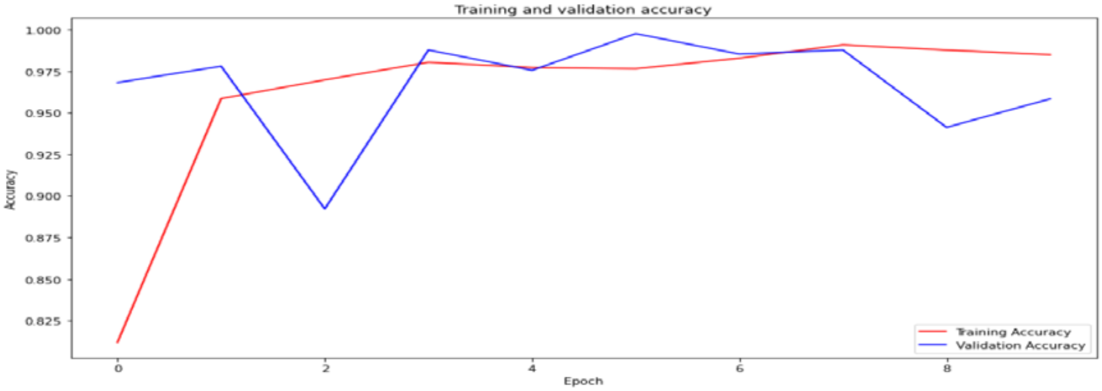


Fig. 10. Training and Validation Accuracy for VVG16.

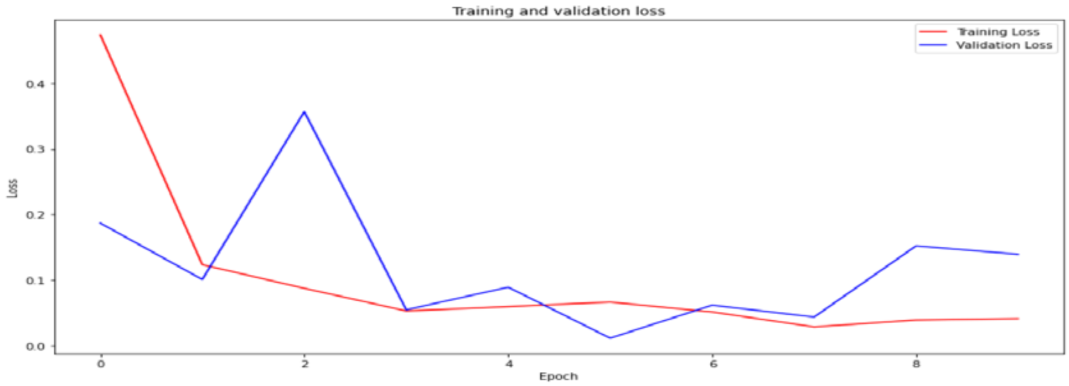


Fig. 11. Training and Validation Loss for VVG16.

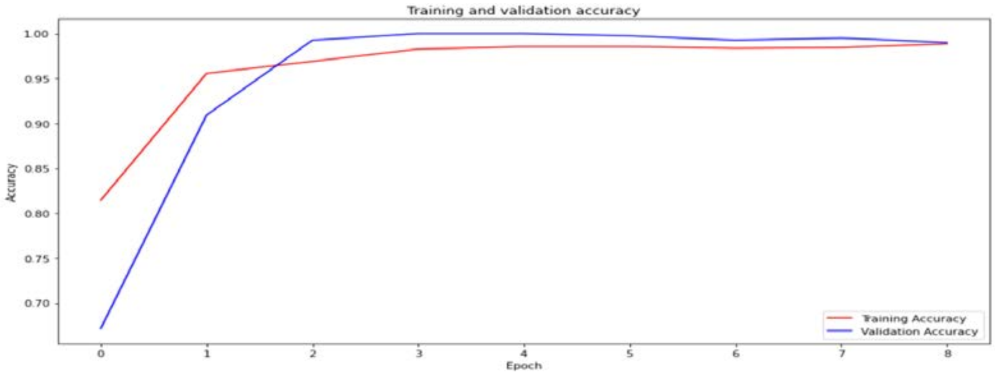


Fig. 12. Training and Validation Accuracy for VVG19.

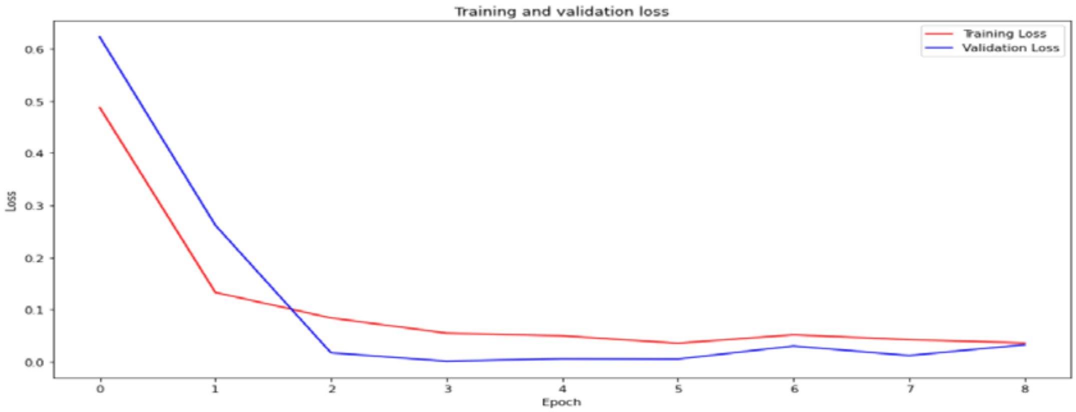


Fig. 13. Training and Validation Loss for VVG19.

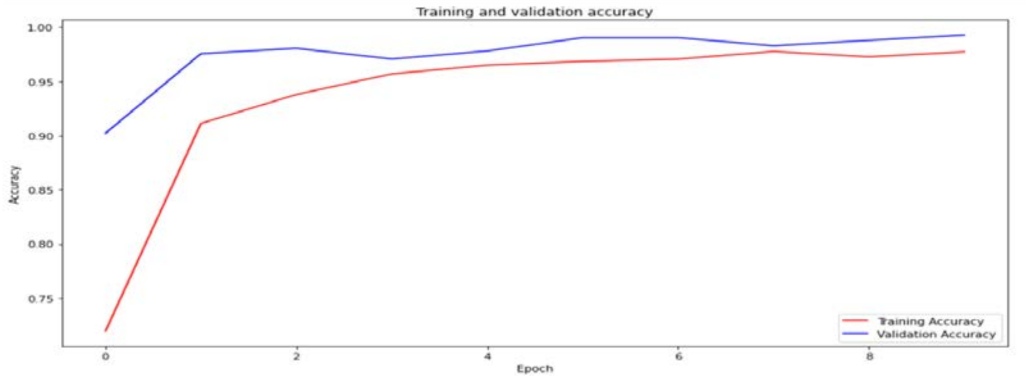


Fig. 14. Training and Validation Accuracy for Xception.

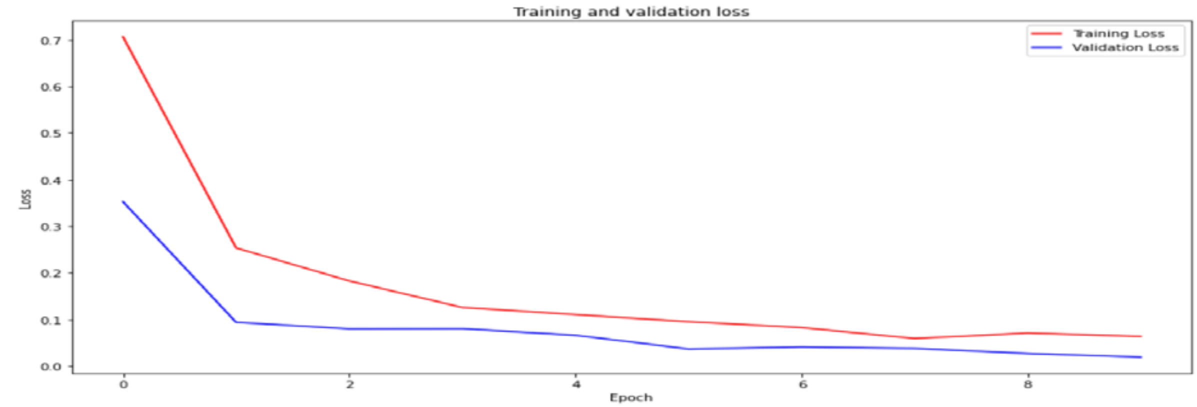


Fig. 15. Training and Validation Loss for Xception.

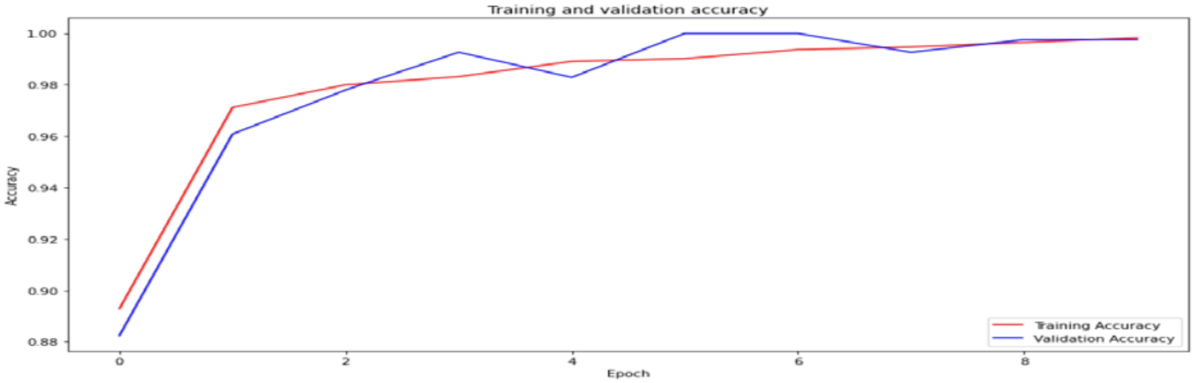


Fig. 16. Training and Validation Accuracy for Proposed Model.

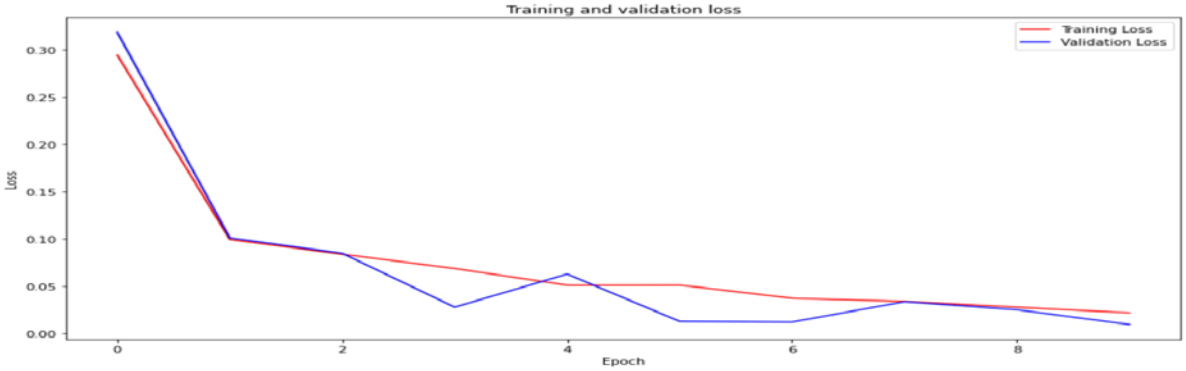


Fig. 17. Training and Validation Loss for Proposed Model.

V. Conclusion

Visual observation for classifying grape leaf diseases can be misleading because of a lack of prior knowledge and similarities among diseases. This paper introduces an automated mechanism for classifying grape leaves as healthy or diseased (e.g., black measles, black rot, and leaf blight) using transfer learning (e.g., VGG16, VGG19, and Xception). This classification has been done by extracting the features from the grape images using different pre-trained networks and then using the ensemble learning method for these networks to enhance the diagnosis accuracy. The novelty of the work lies in the fact that, instead of training each network alone and then concatenating the final results to get better results, all the networks are trained together, allowing the fully connected layer to discover the best ensemble method to concatenate the results of previous networks. The proposed model achieved an accuracy of 99.82% compared to recent models that used grape leaf disease in the training process across the existing online Plant Village dataset. This work demonstrated the value and benefits of using ensemble learning and transfer learning. This study has achieved its objectives to improve the performance of classifying plant diseases and reducing overfitting.

VI. Future Work

Looking ahead, we hope to test the power of our model on more complex datasets and increase the number of categories of agricultural diseases.

References

[1] "Expert meeting on how to feed the world in 2050," Fao.org. [Online]. Available: https://www.fao.org/3/ak971e/ak971e.pdf.

[2] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model," Ecol. Inform., vol. 61, no. 101182, p. 101182, 2021.

[3] S. Ghosh, A. Dasgupta, and A. Swetapadma, "A study on support vector machine based linear and non-linear pattern classification," in 2019 International Conference on Intelligent Sustainable Systems (ICISS), 2019.

[4] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019.

[5] R. Gebbers and V. I. Adamchuk, "Precision agriculture and food security," Science, vol. 327, no. 5967, pp. 828–831, 2010.

[6] Z. J. Wang et al., "CNN explainer: Learning convolutional neural networks with interactive visualization," IEEE Trans. Vis. Comput. Graph., vol. 27, no. 2, pp. 1396–1406, 2021.

[7] Ali, A. (2019). *PlantVillage Dataset* [Data set] [Online]. Available: https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset.

[8] A. Kp and J. Anitha, "Plant disease classification using deep learning," in 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 2021.

[9] Y. Nagaraju, Venkatesh, S. Swetha, and S. Stalin, "Apple and grape leaf diseases classification using transfer learning via fine-tuned classifier," in 2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), 2020.

[10] E. Hirani, V. Magotra, J. Jain, and P. Bide, "Plant disease detection using deep learning," in 2021 6th International Conference for Convergence in Technology (I2CT), 2021.

[11] K. Z. Thet, K. K. Htwe, and M. M. Thein, "Grape leaf diseases classification using convolutional neural network," in 2020 International Conference on Advanced Information Technologies (ICAIT), 2020.

[12] Uky.edu.[Online].Available: https://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1182&context=anr_reports. [Accessed: 06-Apr-2022].

[13] Jim Isleib, Michigan State University Extension, "Signs and symptoms of plant disease: Is it fungal, viral or bacterial?," Field Crops.[Online].Available: https://www.canr.msu.edu/news/signs_and_symptoms_of_plant_disease_is_it_fungal_viral_or_bacterial. [Accessed: 06-Apr-2022].

[14] "Grapes: Diseases and Symptoms," Vikaspedia.in. [Online]. Available:https://vikaspedia.in/agriculture/crop-production/integrated-pest-managment/ipm-for-fruit-crops/ipm-strategies-for-grapes/grapes-diseases-and-symptoms. [Accessed: 06-Apr-2022].

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv [cs.CV], 2014.

[16] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in 2017 International Conference on Engineering and Technology (ICET), 2017.

[17] J. Brownlee, "A gentle introduction to the rectified linear unit (ReLU)," Machine Learning Mastery, 08-Jan-2019. [Online]. Available: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/. [Accessed: 06-Apr-2022].

[18] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," arXiv [cs.NE], 2018.

[19] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv [cs.LG], 2014.

[21] Y. Zhou, X. Wang, M. Zhang, J. Zhu, R. Zheng, and Q. Wu, "MPCE: A maximum probability based cross entropy loss function for neural network classification," IEEE Access, vol. 7, pp. 146331–146341, 2019.

[22] E. Bisong, "Google Colaboratory," in Building Machine Learning and Deep Learning Models on Google Cloud Platform, Berkeley, CA: Apress, 2019, pp. 59–64.

[23] Keras Team, "Getting started," Keras.io. [Online]. Available: https://keras.io/getting_started/. [Accessed: 06-Apr-2022].

An Extractive Text Summarization based on Candidate Summary Sentences using Fuzzy-Decision Tree

Adhika Pramita Widyassari[1], Edy Noersasongko[2], Abdul Syukur[3], Affandy[4]
Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia[1, 2, 3, 4]
Department of Informatic, STT Ronggolawe Cepu, Blora, Indonesia[1]

***Abstract*—This study aims to predict candidate summary sentences in extractive summary using the Fuzzy-Decision Tree method. The fuzzy method is quite superior and the most widely used in extractive summaries, because Fuzzy has advantages in calculations that are not cryptic, so it is able to calculate uncertain possibilities. However, in its implementation, the fuzzy rule generation process is often carried out randomly or based on expert understanding so that it does not represent the distribution of the data. Therefore, in this study, a Decision Tree (DT) technique was added to generate fuzzy rules. From the fuzzy final result, important sentences are obtained that are candidates for summary sentences. The performance of our proposed method was tested on the 2002 DUC dataset in the ROUGE-1 evaluation. The results showed that our method outperformed other methods (baseline and sentence ranking) with an average precision of 0.882498, Recall 0.820443 and F Measure 0.882498 with CI for F1 0.821-0.879 at the 95% confidence level.**

***Keywords*—Text summarization; extractive; fuzzy; decision tree**

I. Introduction

The development of information and communication technology, especially the internet, has an impact on increasing the number of publications of articles on websites or online media which are very useful for decision-making processes and movements regarding everyday life for humans. However, reading the entire text or obtaining relevant information on a particular topic becomes a tedious and time-consuming task. Automated text summarization is recognized as a solution to this problem, because automatic text summaries generate summaries that include all key relevant information quickly without losing the original intent of the text.

There have been many different methods and approaches in the field of text summarization. One approach is abstractive and extractive text summarization. In an abstract summary, sentences generated by summaries are called new sentences, or paraphrases, that use words that are not in the text to generate summaries. Abstract summarization is much more complex and relatively more difficult than extractive summarization. In contrast to abstractive text summarization, the results of the extractive summary consists of fully extracted content, just as the summary result is a sentence or word extracted from the original text [1].

Based on the latest three-year literature, the extractive approach is quite widely used [2] [3]. Several different approaches handled the process of extracting text summarization, one of which is the Frequency-based term weighting approach [4]. The rule-based approach is a study by Naik & Gaonkar [5], which provides the rule-based summarizer with the highest average accuracy, f Measure, and recall values, but has never been tested with broader data contradictions.

In making a summary, the concept of classification can be used by classifying sentences into two groups, namely sentences that are included in the summary (which are important sentences) and sentences that are not included in the summary (sentences that are not important). The fuzzy method has often been used in classification cases and gives good results in both classification and prediction. Fuzzy has the advantage in that its calculations are not rigid (fuzzy), so that it is able to take into account uncertain possibilities [6] [7]. The fuzzy approach is quite widely used in extractive summaries [8]. The fuzzy logic approach is a commonly used method because it can prevent data inconsistencies involving the human role of reviewing sentences and agreeing to select specific sentences to create a summary sentence [9]. The fuzzy system works with different features or multiple inputs from the index. The score for each feature is then passed to the fuzzy inference system as input for later use of IF THEN rules in human knowledge.

Although the fuzzy method is quite superior in extractive summary, in this case, fuzzy has a complexity in terms of determining the rules or basic rules used during inference. Some fuzzy cases use rules obtained by experts, namely humans, while humans can be subjective and can make mistakes. It is feared that this does not reflect the actual data representation. From these problems, a special method is needed for determining the rules of the fuzzy inference system (FIS). Therefore, in this study, the rule to determine candidate summary sentences from FIS will be generated using a Decision Tree. Thus the rules used in the inference engine will represent the actual situation.

In this paper, we propose a text summarizing method that begins by predicting the candidate summary sentences and then compares them with the reference summary results provided by the expert (dataset used is DUC 2002). The proposed summary method is a fuzzy rule-based system for identifying and selecting sentences. To create a fuzzy rule, we use a Decission Tree. Auto-generated text summaries can reduce reading-related cognitive efforts, especially with large amounts of textual information [10]. Contribution to this work is to propose an automatic text summarizing method by predicting candidate summary sentences using fuzzy and decission tree (fuzzy-decission tree) methods in extractive summarizing areas, comparing the proposed method with other methods. The remainder of this paper is organized as follows. Section II describes related work. Section III details this method. Section IV reports experiments for performance evaluation and discussion. Section V discusses conclusions and future work.

II. Related Work

This section analyzes the state of research on automatic text summarization from the last few years using a fuzzy logic approach. The things analyzed include algorithms, datasets, text features, performance, and comparison measures used. Fuzzy logic and NLP are interconnected [11] to solve tasks such as text addition, sentiment analysis, and knowledge representation. This is because fuzzy logic overcomes the problem of inaccuracy and ambiguity of human language by providing a description of the dataset with a linguistic concept defined as a fuzzy set [12]. The following are some of the works that use fuzzy logic in text summarization.

Research conducted by Megala et al., 2014 compares the performance of the fuzzy logic method with the artificial neural network method in summarizing the text. The method is tested by non-automatic evaluation using legal documents. As a result, fuzzy logic is superior in measuring f-measures compared to artificial neural networks, namely 0.46 for fuzzy logic and 0.42 for artificial neural networks [13]. Researchers Megala et al., 2015 again summarized text with fuzzy logic to extract the size, produce a summary and classify segments using Conditional Random Fields (CRF). The evaluation was carried out with legal decision documents, showing that it yielded 0.26 for the f-measure and the method was able to classify all segments in the case [14].

The next research is multi-document summarization using cross-document relationships using fuzzy on news. There are three jobs being carried out, among others: extracting sentence components, performing semantic relationships between text units using Cross-document Structure Theory (CST), and scoring sentences using fuzzy logic. The summary results are obtained from the sentences with the highest ranking. Evaluation reached 0.33 for ROUGE-1 with a 2002 DUC [15]. Summarizing text research with a new model by combining three methods. CLA is used to overcome redundancy, this model uses CLA to reduce redundancy problems, PSO is used to assign feature weights, fuzzy logic is used for sentence assessment. The features used to select important sentences include keywords, sentence length, nouns, thematics, and sentence positions. Method performance excels at f-measure 0.48 when tested on the 2002 DUC dataset [9]. Research on an automatic summary system with a fuzzy approach to extract some features to get important information on student assignment texts. This summary system was tested on a collection of text responses from students to assignments given in a Virtual Learning Environment (VLE). The proposed model is then compared with the method of score, Baseline, sentence and model with ROUGE [16].

Text summary for single document using fuzzy logic approach by extracting sentence features and calculating word scores [17]. Choose a summary sentence by calculating the weight of the sentence and compiling it into a summary. Sentence weights are obtained from calculating word scores and extracting sentence features with fuzzy inference. So that the sentences that stand out are based on fuzzy inference measurements which will later be included in the summary. The method was evaluated using ROUGE-N on the DUC 2002 dataset and then compared with other methods, and the results were superior to the comparison method. The next research is to develop a text summary using fuzzy models in many documents [18]. The fuzzy method is used to handle the uncertainty of feature weights. In this model, the cosine similarity is added to solve the redundancy problem. The evaluation was carried out using ROUGE at DUC 2004. As a result, the proposed method performed higher than the comparison method (Yago Summarizer, TexLexAn, PatSum, ItemSum, and MSSF). In recall measurement, the result is superior, namely 0.1555 on Rouge-2, but lacking in precision when compared to the Patsum method [19].

Extractive summary research that produces an abstract summary. Abstractive summary is obtained by combining the extractive sentence selection process (which uses fuzzy logic) and the long-term two-way short-term memory (Bi-LSTM) method to update the network weights, so it is called the Fuzzy long short-term memory (FLSTM) method. The fuzzy approach is used to get the most important and relevant sentences by extracting information from the document. Important and relevant sentences obtained from the fuzzy extraction method are then used as input for the Bi-LSTM method to produce an abstract summary. The model was then evaluated using ROUGE on the DUC and CNN datasets. The proposed model shows better performance empirically than other comparison methods [20].

Subsequent text summarizing research proposes the integration of two methods, namely the Restricted Boltzmann Machine and Fuzzy logic, hence the name FRBM (Fuzzy Restricted Boltzmann Machine). These two methods have different ways of producing precise summary sentences, but these two methods have something in common, namely, they are unsupervised methods used to summarize text. The summary generated by fuzzy logic is then integrated with the summary generated by the Restricted Boltzmann Machine. The advantage of the FRBM model is that in dealing with noise during training, it is more resistant than RBM [21].

Automatic text summarization uses three different algorithms. It is a two-tailed score for local contextual information (LCIS), key term weighting by sentence, and a fuzzy graph sentence score (FGSS). The LCIS score was used to identify the LCIS, the weighting was used to increase the weight of the important terms, and the fuzzy graph sentence score was used to document the centroid by calculating the appropriate fuzzy graph sentence score. It exhibits superior averages compared to previous studies and requires no training or testing [22].

The next research, summarize unsupervised extractives with fuzzy logic method. Fuzzy logic is used because it is based on natural language that is easy to understand. This summary is built using the Python language. The features used include position, bitcoin, tritoken, cosine similarity, thematic sentence length, numerical data and TF-ISF. The method was tested on the UCI, BBC and DUC 2004 datasets using the ROUGE-1 evaluation. The results conclude that fuzzy logic makes feature extraction sharper and more precise [23]. The next extractive summary research is to summarize the text documents of students' essay assignments using the fuzzy C-Means method. The feature used in this research is the sentence weighting feature. The summary obtained is the sentence with the highest weight in the cluster [24].

In the next research, extractive summaries combine three approaches, namely fuzzy, evolutionary and clustering. The workings of this model begins with clustering, namely grouping sentences according to their similarities. Then extract the significant sentences of each cluster. An evolutionary optimization approach is used to find the optimal weights for text features. Fuzzy inference is used to determine the final score of the sentence. The proposed model was tested on three datasets namely CNN, DUC 2021 and DUC 2022. The results show that the hybrid method produces a good summary [3].

III. Porposed Method

In this section, the proposed method of summarizing extractive techniques using fuzzy and decision tree methods is proposed. Fuzzy method is used to extract features, while decision is used to help fuzzy in making rules (rule based). The proposed model is described in Fig. 1 and is broken down into five main steps, as follows:

A. Preprocessing

Preprocessing is the initial stage for text summarization. Some of the pre-processing steps needed in this research are as follows: Removing punctuation and special words that are not used. Segmentation, the process of separating text into sentence units. Tokenization is the process of separating words from each sentence. Stop word removal, removing a collection of unused words. And steamming is the process of converting each word to its basic form by removing prefixes, affixes and suffixes.

B. Features used

At this stage, describe the extracted features. Feature extraction is used to get important sentences from text sources that have been preprocessed before. This feature ensures the importance of each sentence that goes into the summary. So that the summary contains sentences with high scores. Therefore, selecting the right features can have a big impact on the quality of the summary. In this study, used seven extracted features for each sentence in the input data. The seven features are:

1) Sentence position: The key concept here is that sentences that appear at the beginning or end of the input text are considered more important than other sentences. So for the initial and final sentences, initialize 1 and those that are not the initial and final sentences are initialized 0 [25].

Score (S_i) *= 1 for First and Last sentence,*

0 for other sentences (1)

2) Sentence lenght (in document): The sentence length feature in the document is used to filter out short sentences such as the author's name, address and date that might be found in news documents. Sentences that are too short are not expected to be part of the summary [5]. So that in this feature normalization of sentence length is carried out, namely the ratio of the number of words that appear in the sentence to the number of words that appear in the longest sentence of the document [25].

$$Score\ (Si) = \frac{No.of\ a\ word\ occurring\ in\ S_i}{No.Word\ Occurring\ in\ longest\ Sentence\ from\ document} \tag{2}$$

3) Sentence length (in paragraph): In this feature, short sentences may not represent the topic of the document because the words contained in it are few. Thus, long and short sentences are given low scores. The value of sentence length in paragraphs is calculated based on the equation: [26]

$$Score\ (Si) = \frac{No.of\ a\ word\ occurring\ in\ S_i}{No.Word\ Occurring\ in\ longest\ Sentence\ from\ paragraph} \tag{3}$$

4) Thematic word: This feature is quite important in text summarization because terms that appear frequently in a document may be related to the topic. The thematic word count indicates the words with the maximum relativity possible [25]. After getting the thematic words in the text, then look for the ratio of the number of words in the sentence that appears in the thematic word to the length of the sentence (the number of words in the sentence).

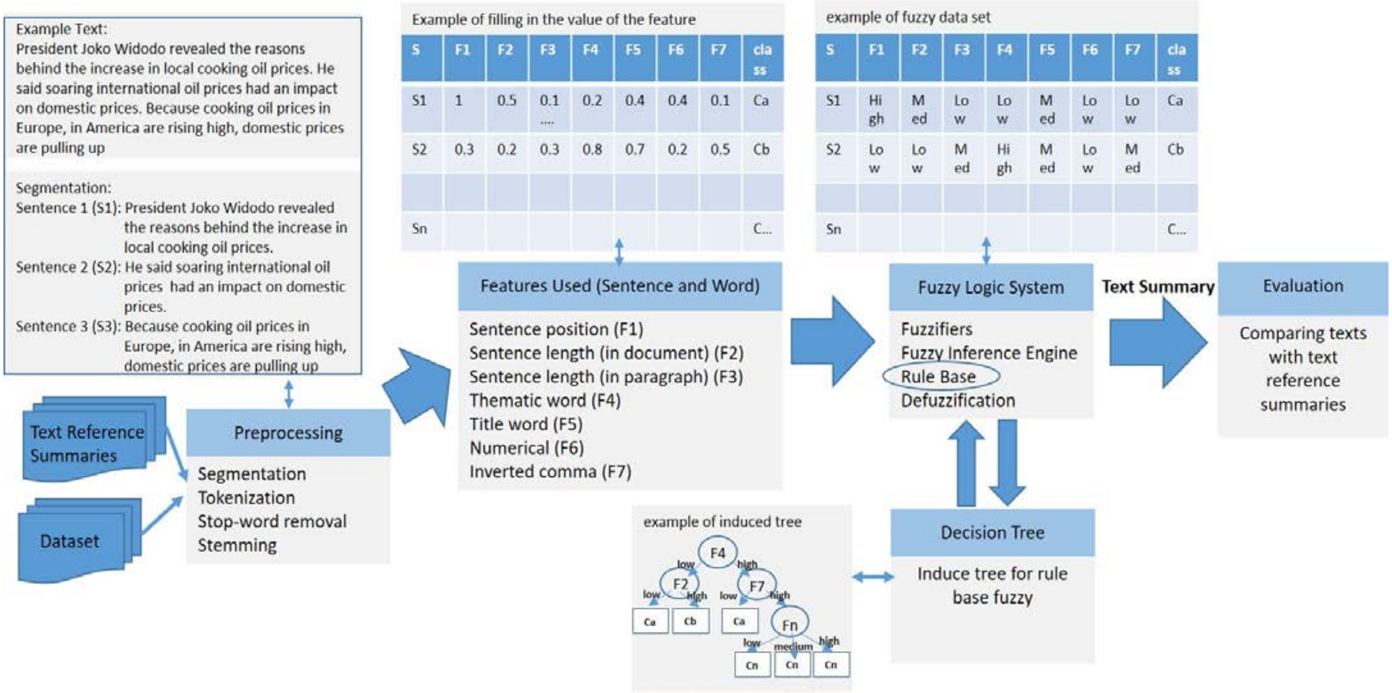
$$Score\ (S_i) = \frac{No.Thematic\ word\ in\ S_i}{Length\ (Si)} \tag{4}$$

5) Title word: This feature is used to find the number of words in the title that appear in the sentence. Words in sentences that also appear in the title is given a high score [6]. Calculate score for this feature which is the ratio of the number of words in a sentence that appears in the title to the length in that sentence (the number of words in that sentence).

$$Score\ (S_i) = \frac{No.Title\ word\ in\ S_i}{No.word\ occuring\ in\ S_i} \tag{5}$$

6) Numerical: In the summary, text or sentences containing numbers are considered informative. So in this feature, sentences with a lot of numerical data need to be considered to be included in the summary. The way to calculate it is to calculate the ratio of the number of numeric data in a sentence divided by the length of the sentence [21].

$$Score\ (S_i) = \frac{No.Numerical\ data\ in\ S_i}{\text{No.word occuring in } S_i} \tag{6}$$



Example Text:
President Joko Widodo revealed the reasons behind the increase in local cooking oil prices. He said soaring international oil prices had an impact on domestic prices. Because cooking oil prices in Europe, in America are rising high, domestic prices are pulling up

Segmentation:
Sentence 1 (S1): President Joko Widodo revealed the reasons behind the increase in local cooking oil prices.
Sentence 2 (S2): He said soaring international oil prices had an impact on domestic prices.
Sentence 3 (S3): Because cooking oil prices in Europe, in America are rising high, domestic prices are pulling up

Example of filling in the value of the feature

S	F1	F2	F3	F4	F5	F6	F7	class
S1	1	0.5	0.1	0.2	0.4	0.4	0.1	Ca
S2	0.3	0.2	0.3	0.8	0.7	0.2	0.5	Cb
Sn								C...

example of fuzzy data set

S	F1	F2	F3	F4	F5	F6	F7	class
S1	High	Med	Low	Low	Med	Low	Low	Ca
S2	Low	Low	Med	High	Med	Low	Med	Cb
Sn								C...

Features Used (Sentence and Word)
Sentence position (F1)
Sentence length (in document) (F2)
Sentence length (in paragraph) (F3)
Thematic word (F4)
Title word (F5)
Numerical (F6)
Inverted comma (F7)

Fig. 1. The Proposed Method of Summarizing Extractive Techniques using Fuzzy and Decision Tree.

7) Inverted comma: Inverted commas usually indicate direct conversation, title or name, and also contain important information [26]. The inverted comma is calculated using the following equation.

$$Score\ (S_i) = \frac{No.Inverted\ comma\ in\ S_i}{No.word\ occuring\ in\ S_i} \quad (7)$$

C. Fuzzy Logic System

Fuzzy logic has often been used for various applications because fuzzy logic is easy to understand, flexible and tolerant of inaccurate data. One of the characteristics of fuzzy logic is the use of verbal instructions described in fuzzy sets and rules. The way the fuzzy logic system works there are four parts as follows:

1) Fuzzifiers: The first process in the Fuzzy system is to transform raw crisp values into membership values through membership functions. This means that the membership function for each fuzzy set must be determined first. In this section, the input is the value of text summarizing features (in the form of numbers), which by using the membership function will be converted into lingiustic. The membership function used for this summary model is the Triangular Membership Function (TMF). Where each feature has three fuzzy sets, namely: high, medium and low.

2) Fuzzy inference engine: This section is the main part of fuzzy logic. Here will be calculated formulas so as to produce output. There are two things that serve as a reference for this calculation, namely the membership function in the previous section and also the fuzzy rule. So that fuzzy input is needed from the fuzzifier in making decisions based on rules. For our proposed summary model, the inference used is Mamdani fuzzy inference (FIS). Due to its simple and most common min-max operating structure it is used in many applications. In addition, Mamdani is more suitable for text summarization systems because it can capture expert knowledge which allows us to describe abilities in a more insightful and more human-like way. To process it using the help of MATLAB.

3) Rule base: The rule design process is an important part in the fuzzy classification algorithm. In this study, to help activate the rule, the decision tree method was used in designing the if-then rules. From the results of making the decision tree, 33 if-then rules are obtained. To see a more detailed explanation of the decision tree method, see section 3, part d.

4) Defuzzification: converting linguistic inference results back into sharp outputs. In this study, the centroid defuzzification method was chosen for us to use. This method is the default method where it works by returning the center of the area under the curve.

D. Decision Tree (for Rule Fuzzy)

Decision tree is one of the most famous studies to describe the decision-making process based on existing knowledge. Each branch of the decision tree can be converted into a decision rule, and all these decision rules can generate a decision rule base (Mu et al., 2019). Therefore, in this research, each feature value that is input in fuzzy where the linguistic fuzzifier process has been made instead of the membership function of each feature, will be used as training data for the decision tree algorithm to produce a decision model. To produce a decision tree, the C4.5 algorithm is used to process the training data. This stage begins with calculating the entropy value that will be used to calculate the gain value for each feature. The feature with the highest gain value will then be set as root. The formulas for calculating entropy and gain are shown in Equation (8) and Equation (9). The step of calculating entropy and gain for each feature is repeated continuously until all features are partitioned.

$$Entropy(S) = \sum_{n=1}^{c} - {p_i}^2 \, log \, p_i \qquad (8)$$

$$Gain(S, A) = Entropy(S) - \sum\nolimits_{v \in Values(A)} \frac{|S_v|}{S} Entropy(S_v) \quad (9)$$

From the decision tree that has been made, 33 if-then rules are obtained which will be used in the fuzzy process. Fig. 2 below shows one of the rules formed from the results of the decision tree.

IF (Thematic is medium) and (Sentence length in a paragraph is high) and (Sentence length in a document is medium) and (Sentence position is high) and (Numerical is low) and (Inverted Comma is low) and (Tittle word is medium) THEN (sentence is Yes)

Fig. 2. Sample IF THEN RULE Results from Decision Tree.

E. *Evaluation*

In the evaluation of this research, the results of the text summary from the system will be compared with the reference summary from the expert (human). In the evaluation used the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) which was used to measure the similarity between system generated summaries and human reference summaries (Lin & Hovy, 2003). The ROUGE evaluation produces three measures, namely: recall, precision and f-measure. The measure is calculated by counting the words that overlap between the computer-generated summary and the human-generated ideal summary. Precision is the number of n-grams that appear together in the system summary and reference summary divided by the total number of n-gram reference summaries. Whereas Recall is the number of n-grams that appear together in the system summary and the reference summary divided by the total number of n-grams in the system summary. Precision and recall ranged from 0 to 1. When the precision score was 1, all n-grams in the text summary were in the reference summary. F-Measure is a combination of precision and recall, which is a weighted harmonic average of precision and recall. The study of Steinberger & Jezek, 2009 showed that automated evaluation using the unigram version of ROUGE-N, namely ROUGE-1 correlated well with human evaluations based on various statistics. Therefore, this study uses the evaluation of the system summary results with ROUGE-1.

IV. Experiment and Result

This section describes the evaluation of the performance results of the Fuzzy Decision Tree method. Shows a comparison of our proposed Method with other methods. The Baseline method and the Sentence method are also implemented to be used in the evaluation process as a comparison to the Fuzzy-Decision Tree method.

A. *Compared Methods*

1) Fuzzy-decision tree: In our Fuzzy-Decision Tree method, the first start by setting the fuzzy set in three input variables: high, medium, low (for each feature). The selected sentences must represent an informativeness or indicate the level of importance of a sentence classified in the output variable as YES/NO. To assist our work in designing fuzzy models, used the Fuzzy Logic Toolbox in MATLAB. The type of inference, used is Mamdani. To help create an IF THEN rule, used the Decision Tree method.

2) Baseline: The baseline system is the basic information collected before a program starts. For comparison if using DUC 2002, use Baseline-1 DUC 2002. Baseline-1 is the first 100 words from the beginning of the document as determined by DUC 2002 https://www-nlpir.nist.gov/projects/duc/duc2002/baselines.html [25].

3) Sentence ranking: Sentence method selects sentences based on word frequency. First of all choose fetch keywords. After that it calculates the frequency of each keyword such as how often it appears from the maximum frequency this keyword is taken and calculates the number of frequency weights. And the last one is extracting high-frequency sentences [27].

B. *Dataset*

In this experiment, 10 data were taken in the form of text documents from the DUC 2002 dataset. So that the total words in this experiment were 3803, the total sentences were 242 sentences. Each text document contains an average of 380 words and an average of 24 sentences. Data for evaluation, 10 reference summary documents have been provided by language course experts based on the desired key concepts. The total data consists of 2122 words and 98 sentences. Each reference text document contains an average of 212 words and eight sentences. As an evaluation of the comparison between the summary data generated by the system and the reference summary data, used n-gram statistics. Measurement with n-gram ROUGE has a 95% confidence level so that it is highly correlated with human judgment.

C. *Result and Discussion*

For evaluation of summary texts, this study uses a set of ROUGE metrics that have become the standard for automatic summary evaluation. Evaluation is done by comparing the results of the summary of the system with the results of the summary of human references. To compare summaries, n-grams are used. ROUGE-I is consistently highly correlated with human judgment and has high recall and significance test precision with manual evaluation results. So in the experiment of summarizing the text, the ROUGE-1 measurement was used. Table I shows the comparison of the summary results of the proposed method, namely the Fuzzy-Decision Tree with the summary results of the baseline and summary results of the sentence method from the 2002 DUC collection.

TABLE I.

PERFORMANCE COMPARISON

Document	Fuzzy-Decision Tree			Baseline			Sentence Ranking [27]		
	R	P	F1	R	P	F1	R	P	F1
AP880911-0016	0.75	0.825	0.78571	0.33333	0.5	0.4	0.72727	0.65306	0.68817
AP900621-0192	0.82609	0.83333	0.82969	0.36522	0.36207	0.36364	0.69565	0.65574	0.67511
AP900621-0186	0.85586	0.94059	0.89623	0.43243	0.61538	0.50793	0.7027	0.77228	0.73585
AP880821-0008	0.85938	0.9322	0.89431	0.34375	0.46809	0.3964	0.78906	0.71631	0.75093
AP880228-0013	0.76866	0.88793	0.824	0.40299	0.53465	0.45958	0.87313	0.84783	0.86029
AP880508-0070	0.808	0.93519	0.86695	0.4	0.65789	0.49751	0.736	0.77966	0.7572
AP881025-0196	0.74803	0.84071	0.79167	0.33071	0.53165	0.40777	0.65354	0.72807	0.68879
AP880808-0040	0.89011	0.84375	0.86631	0.51648	0.63514	0.5697	0.65934	0.61224	0.63492
AP900103-0077	0.86957	0.90909	0.88889	0.48913	0.6	0.53892	0.61957	0.75	0.67857
AP880914-0027	0.82873	0.87719	0.85227	0.34807	0.55752	0.42857	0.80663	0.90123	0.85131
Average	**0.820443**	**0.882498**	**0.849603**	**0.396211**	**0.546239**	**0.457002**	**0.726289**	**0.741642**	**0.732114**

R = Recall, P = Precision, F1= F Measure

The application of the ROUGE-1 metric resulted in the performance of the tested method on 10 data taken from the DUC 2002 dataset which is presented in Table I. From the table it is shown that the Fuzzy-Decision Tree method has superior performance in all documents tested except for the AP880228-0013 document, where the sentence method outperforms the Fuzzy-Decision Tree in the evaluation of recall 0.87313 and F1 0.86029.

Table II highlights the average performance results of F-Measure, Recall and Precision produced by the Fuzzy-Decision Tree method, the Baseline method and the sentence ranking method. From the table, it shows that the Fuzzy-Decision method has the best average, namely the average F Measure is 0.882498, the average precision is 0.882498 and the average recall of 0.820443 with CI for F1 0.821-0.879. Followed by the performance of the sentence ranking method which is close to the Fuzzy-Decision Tree method, namely the average F-Measure is 0.732114, the average precision is 0.741642 the recall average is 0.726289 with a CI for F1 of 0.678-0.786. Considering the dataset used is DUC 2002, which is news text, the results are reasonable, because the performance for Recall and precision will be higher if the text used is short text.

The results show that the Fuzzy-Decision Tree performance is significantly better than baseline summary and Sentence ranking. Then compared the performance of the Fuzzy-Decision Tree summary and other summarizers by checking for precision and recall. In this case, the best precision and recall of Fig. 3 and Fig. 4, provide strong evidence of its feasibility in text summarization applications.

Fig. 5 describes the results of the method based on the performance confidence interval (CI). The CI and f measures of the method indicate that the systems cannot be considered equal. The performance of the Sentence ranking method is almost close to the Fuzzy-Decision Tree method, while the Baseline method remains the farthest. This is because the Baseline method only selects the first sentence of the original text. Poor performance of Baseline method due to its simplicity compared to other methods. The sentence ranking method is close to the fuzzy decision tree because the sentence ranking method is based on word frequency, where word frequency represents most of the summary content. And this shows that there is a correlation between one of the features in the Fuzzy-Decision Tree method, namely the thematic word feature because it counts themes that often appear in a document that may be related to the topic.

TABLE II. SUMMARY OF PERFORMANCE COMPARISON

Method	R	P	F1	C1 for F1
Fuzzy-Decision Tree	**0.820443**	**0.882498**	**0.849603**	**0.821-0.879**
Baseline	0.396211	0.546239	0.457002	0.408-0.506
Sentence	0.726289	0.741642	0.732114	0.678-0.786

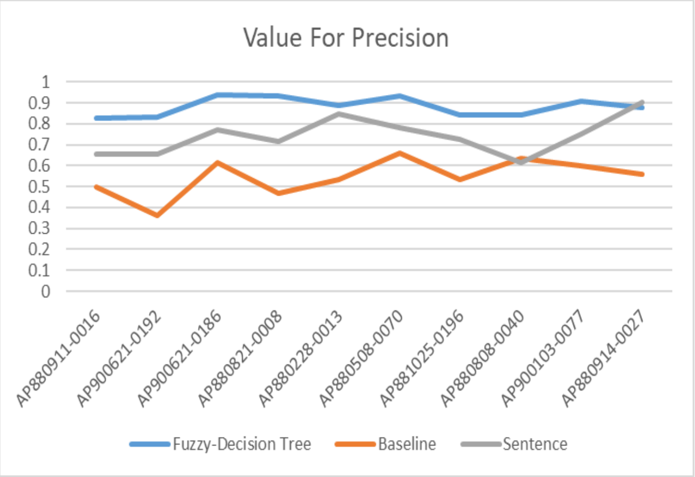


Fig. 3. Precision Result Comparison.

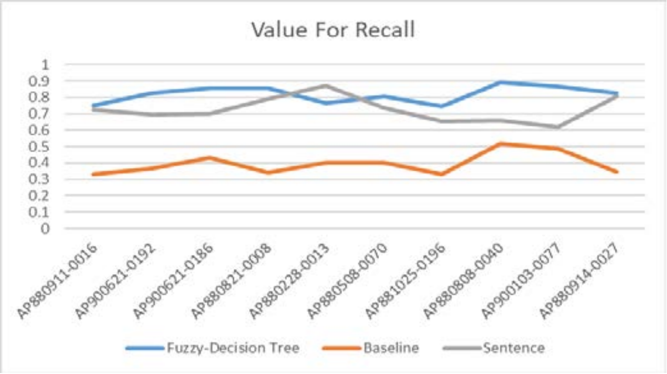


Fig. 4. Recall Result Comparison.

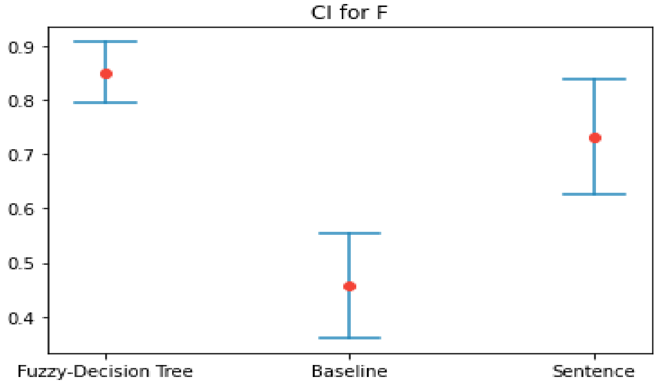


Fig. 5. Confidence Interval (CI) for F1.

V. Conclusion and Future Work

In this study a text summarization is proposed that starts by predicting candidate summary sentences and then compares it with the results of a human-given reference summary. The summary method proposed is a system with a fuzzy approach that identifies and selects important sentences based on important features. Important features in this study include: sentence position, sentence length in the document, sentence length in paragraphs, thematic words, title words, numeric data and inverted commas. Fuzzy logic is used because it is believed to be able to handle uncertain information such as language ambiguity. To optimize the creation of a fuzzy rule base, a decision tree method is added so that the built rules reflect the actual data representation. Contributions in this paper include: 1. Combination method between fuzzy logic and decision tree (fuzzy-decision tree) which has never been applied to extractive summary research before; 2. Comparison of the proposed method (fuzzy-decision tree) with other methods (baseline and sentence ranking).

Evaluation of the proposed method shows that our method outperforms other methods included in the comparison method, namely baseline and sentence ranking. By evaluating ROUGE-1, our method excels with mean precision 0.882498, mean drawdown 0.820443 and F Measure 0.882498 with CI for F1 0.821-0.879 at 95% confidence level tested in single-document test data. However, the performance of the method has not been tested on multi-documents.

For our next work that is part of this research is to add other important features that have not been used in our proposed summary, such as the similarity feature between sentences thereby reducing sentence redundancy in the summary, and the word frequency feature for making summaries. The results are more informative comparing with other more diverse summary methods and adding focus to summary results not only on quality but also on more diverse summary quantity such as 20%, 30% or 40% summary.

References

[1] M. Gambhir and V. Gupta, “Recent automatic text summarization techniques: a survey,” Artif. Intell. Rev., vol. 47, no. 1, pp. 1–66, 2017.

[2] P. Ren et al., “Sentence Relations for Extractive Summarization with Deep Neural Networks,” ACM Trans. Inf. Syst., vol. 36, no. 4, 2018.

[3] P. Verma, A. Verma, and S. Pal, “An approach for extractive text summarization using fuzzy evolutionary and clustering algorithms,” Appl. Soft Comput., vol. 120, 2022.

[4] K. U. Manjari, S. Rousha, D. Sumanth, and J. Sirisha Devi, “Extractive Text Summarization from Web pages using Selenium and TF-IDF algorithm,” in Proceedings of the 4th International Conference on Trends in Electronics and Informatics, ICOEI 2020, 2020, no. June, pp. 648–652.

[5] S. S. Naik and M. N. Gaonkar, “Extractive text summarization by feature-based sentence extraction using rule-based concept,” in RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings, 2017, vol. 2018–Janua, pp. 1364–1368.

[6] C. Dumitrescu, P. Ciotirnae, and C. Vizitiu, “Fuzzy logic for intelligent control system using soft computing applications,” Sensors, vol. 21, no. 8, pp. 1–33, 2021.

[7] Y. Yin, Y. Sheng, Y. He, and J. Qin, “Modeling vague spatiotemporal objects based on interval type-2 fuzzy sets,” Int. J. Geogr. Inf. Sci., vol. 36, no. 6, pp. 1258–1273, 2022.

[8] H. Van Lierde and T. W. S. Chow, “Learning with fuzzy hypergraphs : A topical approach to query-oriented text summarization,” Inf. Sci. (Ny)., vol. 496, pp. 212–224, 2019.

[9] R. Abbasi-ghalehtaki, H. Khotanlou, and M. Esmaeilpour, “Fuzzy evolutionary cellular learning automata model for text summarization,” Swarm Evol. Comput., pp. 1–16, 2016.

[10] S. Saraswathi, M. Hemamalini, S. Janani, and V. Priyadharshini, “Multi-document Summarization for Query Answering E-learning System,” Int. J. Comput. Sci. Eng., vol. 3, no. 3, pp. 1147–1154, 2011.

[11] E. Cambria and B. White, “Jumping NLP curves: A review of natural language processing research,” IEEE Comput. Intell. Mag., vol. 9, no. 2, pp. 48–57, 2014.

[12] A. Ramos, J. M. Alonso, E. Reiter, K. van Deemter, and A. Gatt, “Fuzzy-based language grounding of geographical references: From writers to readers,” Int. J. Comput. Intell. Syst., vol. 12, no. 2, pp. 970–983, 2019.

[13] S. S. Megala, A. Kavitha, and A. Marimuthu, “Feature Extraction Based Legal Document Summarization,” Int. J. Adv. Res. Comput. Sci. Manag. Stud., vol. 2, no. 12, pp. 346–352, 2014.

[14] S. Santhana Megala, D. A. Kavitha, and A. Marimuthu, “Text Summarization System using Fuzzy Logic and Conditional Random Field Algorithm,” Int. J. Innov. Res. Comput. Sci. Eng., no. 1, pp. 2394–6364, 2015.

[15] Y. J. Kumar, N. Salim, A. Abuobieda, and A. T. Albaham, “Multi document summarization based on news components using fuzzy cross-document relations,” Appl. Soft Comput. J., vol. 21, pp. 265–279, 2014.

[16] F. B. Goularte, N. Silvia Modesto, R. Fileto, and H. Saggion, “A Text Summarization Method Based on Fuzzy Rules and Applicable to Automated Assessment,” Expert Syst. Appl., vol. 115, pp. 264–275, 2019.

[17] D. Patel and H. Chhinkaniwala, “Fuzzy logic-based single document summarisation with improved sentence scoring technique,” Int. J. Knowl. Eng. Data Min., vol. 5, no. July 2009, pp. 125–138, 2018.

[18] D. Patel, S. Shah, and H. Chhinkaniwala, “Fuzzy logic based multi document summarization with improved sentence scoring and redundancy removal technique,” Expert Syst. Appl., vol. 134, pp. 167–177, 2019.

[19] J. Qiang, P. Chen, W. Ding, F. Xie, and X. Wu, “Multi-document Summarization using Closed Patterns,” Knowledge-Based Syst., 2016.

[20] R. Bhargava and Y. Sharma, “Deep Extractive Text Summarization,” Procedia Comput. Sci., vol. 167, no. 2019, pp. 138–146, 2020.

[21] B. Sharma, M. Tomer, and K. Kriti, “Extractive text summarization using F-RBM,” J. Stat. Manag. Syst., vol. 23, no. 6, pp. 1093–1104, 2020.

[22] T. Vetriselvi and N. P. Gopalan, “An improved key term weightage algorithm for text summarization using local context information and fuzzy graph sentence score,” J. Ambient Intell. Humaniz. Comput., 2020.

[23] A. Khanna, D. Gupta, V. Snasel, and J. Platos, “Automatic Text Summarization Using Fuzzy Extraction,” in Advances in Intelligent Systems and Computing, International Conference on Innovative Computing and Communication, 2020, vol. 1087, pp. 395–405.

[24] I. M. Suwija Putra, Y. Adiwinata, D. P. Singgih Putri, and N. P. Sutramiani, “Extractive Text Summarization of Student Essay Assignment Using Sentence Weight Features and Fuzzy C-Means,” Int. J. Artif. Intell. Res., vol. 5, no. 1, pp. 13–24, 2021.

[25] L. Suanmali, M. S. Binwahlan, and N. Salim, “Sentence features fusion for text summarization using fuzzy logic,” Proc. - 2009 9th Int. Conf. Hybrid Intell. Syst. HIS 2009, vol. 1, pp. 142–146, 2009.

[26] N. Desai and P. Shah, “Automatic Text Summarization Using Supervised Machine Learning Technique for Hindi Langauge,” Int. J. Res. Eng. Technol., vol. 05, no. 06, pp. 361–367, 2016.

[27] J. N. Madhuri and R. Ganesh Kumar, “Extractive Text Summarization Using Sentence Ranking,” in 2019 International Conference on Data Science and Communication, IconDSC 2019, 2019, pp. 1–3.

Design and Implementation of ML Model for Early Diagnosis of Parkinson’s Disease using Gait Data Analysis in IoT Environment

Navita Mehra[1]
Ph.D Scholar
Dept. of Comp. Sci. & App
Maharshi Dayanand University, Rohtak, India

Pooja Mittal[2]
Assistant Professor
Dept. of Comp. Sci. & App
Maharshi Dayanand University, Rohtak, India

***Abstract*—Parkinson’s disease (PD) is the world’s second most neurodegenerative disorder that results in a steady loss of movement. The symptoms in patients occur slowly with the passage of time are and very hard to identify in its initial stage. So, early diagnosis of PD is the foremost need for timely treatment to people. The introduction of smart technologies like the Internet of Things (IoT) and wearable sensors in the healthcare domain offers a smart way of identifying the symptoms of PD patients. In which smart sensors are worn on the patient’s body which continuously monitor the symptoms in patients and track their possible health status. The major objective of this work is to propose a machine learning-based healthcare model that best classifies the subjects into healthy and Parkinson's patients by extracting the most important features. A step regression-based feature selection method is followed to improve the classification of PD. A Shapiro Wilk test is adopted to check the normality of the gait dataset. This model is implemented on three publicly available Parkinson’s datasets collected from three different studies available on Psyionet. All these data sets contain VGRF recordings obtained from eight different sensors placed under each foot. Experimentation is done on the Jupyter notebook by utilizing Python as a programming language. Experimental results revealed that our proposed model with effective pre-processing, feature extraction, and feature selection method achieved the highest accuracy result of 95.54%, 98.80%, and 94.52% respectively when applied to three datasets. Our research inducts knowledge about significant characteristics of a patient suffering from PD and may help to diagnose and cure at an early stage.**

Keywords*—*Internet of things (IoT); sensors; parkinson's disease (PD); machine learning (ML); vertical ground reaction force (VGRF)

I. Introduction

The old age people of today's society suffer from a large number of neurodegenerative disorder diseases like Alzheimer’s disease, dementia and Parkinson's Disease (PD), etc. PD is the second most affecting neurodegenerative disease after Alzheimer's disease affecting people worldwide. A report generated by the Parkinson's Foundation states that 10 million people live with Parkinson's disease (PD) worldwide and among these, approximately one million people are from the United States (US). This report also states that men are 1.5 times more affected by PD as compared to women [1]. 1.04 million people were diagnosed with PD in the US in 2017 and it is estimated to be 1.6 million by 2037 [2]. The progression of PD differs from one person to another person and there does not exist any standard test for the diagnosis of PD and detection is done only based on observations of its symptoms. The clinical demonstrations of PD mainly include tremors in hands, slowing movement, limb rigidity, altered taste in smell, posture instability, etc. [3]. The symptoms in patients occur slowly with time so it's very hard to detect or diagnose this. PD mainly arises due to the degradation of dopamine cells in the brain which control the movement of limbs in the body. By the time symptom appears 50 to 80% of dopamine neurons have already died [4].

Recent discoveries reveal that analysis of gait patterns can be considered a better approach for the detection of neurodegenerative diseases like PD [5] [6]. In the past few years, the invention of smart motion analysis systems and sensors-based motion-capturing devices offers an opportunity for researchers to work on more advanced gait analysis techniques. Smart wireless sensor devices like smart watches, smart bracelets, stand-alone video cameras, smartphones, insole sensors, and other portable devices provide the easiest way of capturing the motion and movement disorders among patients with PD [7]. Major approaches utilized for analysis of gait dataset include smart motion capturing cameras, inertial measurement unit (IMU) based sensors to discover sharp motion and body force in a particular direction [8], plantar pressure determined by utilizing planter sensors, some force subtle platforms for measuring VGRF, High-resolution Electromyography (EMG) devices to capture muscles actions [9]. The neuroimaging approach involves expensive optical cameras and force platforms, while foot-worn sensors offer a reliable, fast, simple, and reasonable gait analysis approach. However, gait analysis faces lots of challenges like high data dimensionality, nonlinear data dependency, and complex correlation.

In recent years, ML techniques have revealed the impressive capability to support clinicians not in identifying the existence of PD but also supports in categorizing the states of PD on the basis of the motor indicators of the subjects [10] [11].

For the detection of PD, ML techniques have been applied to different kinds of measured data like handwritten patterns

[12-14], speech data, neuroimaging data [15], smell identification data [16], spontaneous cardiovascular oscillations [17], and gait data [18]. Many researchers are now working on the early, precise, and timely detection of PD, particularly when ML techniques are applied to learn major strategies. UPDRS and H&Y are two important rating scales utilized for monitoring the progression of PD [19] [20]. H&Y scale is the most commonly utilized rating scale for effective validation of severity level according to functional disability. Therefore, the major objective of the current study was to develop an ML model that could help physicians to diagnose PD by utilizing gait data generated from IoT-based wearable sensors. After that performance of that model is analyzed by using different performance evaluation measures. The major contribution of this work is as follows:

1) This work provides a significant way of detecting/ dosing the PD in patients using an IoT environment.

2) This work proposed a PD diagnosis model by utilizing effective feature extraction and selection method.

3) This work also adopts various performance evaluation metrics to predict healthy and PD patients with the help of collected data patterns.

4) The work also focuses on the comparison of the proposed model on the basis of selected and unselected features set in classifying healthy and PD patients in an IoT environment.

This paper is organized as follows: Section II describes the major causes, symptoms, and measurable indicators of PD. Section III introduces materials and methods utilized for our experimental purpose. Section IV presents different performance metrics utilized for evaluating the performance of the proposed model. Section V describes the result obtained through the implementation of our model. Conclusions are drawn in Section VI. In last, limitations and future directions are described in Section VII.

II. Outline About Major Causes, Symptoms and Measurable Indicators of PD

Parkinson’s disease occurs due to brain disorder and is a progressive neurodegenerative disease that results in inadvertent or non-controllable movements, such as shaking, toughness, and difficulty with steadiness and coordination as shown in Fig. 1.

Categorization of major symptoms of PD is described Major symptoms of PD are described below:

1) Motor symptoms (movement-based symptoms)

- slowed movement (bradykinesia) means muscles weakness,
- tremor means muscles are at rest
- rigidity or stiffness
- Unbalanced postures etc.

2) Non-Motor symptoms (non-movement-based symptoms)

- depression
- loss of smell sense
- sleep disorder
- Inconvenience in thinking and focusing etc.

Even though, PD is incurable because the symptoms of PD usually start steadily and become worse over time. So, early detection of PD can assist to follow proper medication/surgical treatment so that the symptoms can be alleviated. H & Y and UPDRS are two generally utilized clinical ranking scales for monitoring the progression of PD [22] [23]. The former rating scale considers only motor symptoms while the latter assesses both motor and non-motor symptoms.

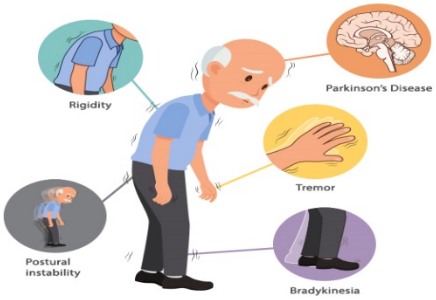


Fig. 1. Parkinson ’s Disease Symptoms Appearance [21].

III. IoT Approach Toward PD Diagnosis

In the traditional medical scenario, diagnosis of PD is a very hard task because it involves proper tracking of the patient's tasks throughout the day. Because there can be a gradual variation in symptoms throughout the day. But patients have no proper sources to track their activities and may lose the most important observations. Another way that the patient moves to the clinician for proper assessment of their postural steadiness and rigidity level which requires transportation cost and time.

The advancement of various smart sensor-based technologies and their integration into the healthcare system reduces the pressure of treatment of various neurological diseases like Parkinson’s disease. “Kevin Ashton” 1999 introduced the concept of the Internet of Things (IoT) which offers connectivity among various wearable sensors and internet platforms that result in amazing remote monitoring services for patients [24]. So, with the introduction of IoT tracking the major symptoms of PD becomes easy [25]. Symptoms collection is possible by just simply placing a few small wireless sensors on the patient body and remotely monitoring the symptoms and details of the patient. So, quality of life can be improved by importing new smart technologies.

Fig. 2 illustrates how a PD patient is treated in an IoT environment while performing any activity. Smart sensors were worn on patients' bodies while they performed their activities. The recording of smart sensors was continuously transferred to the system-based storage or cloud using various transmitting technologies like Bluetooth, Wi-Fi, etc., and after that data is analyzed by utilizing various machine intelligence techniques.

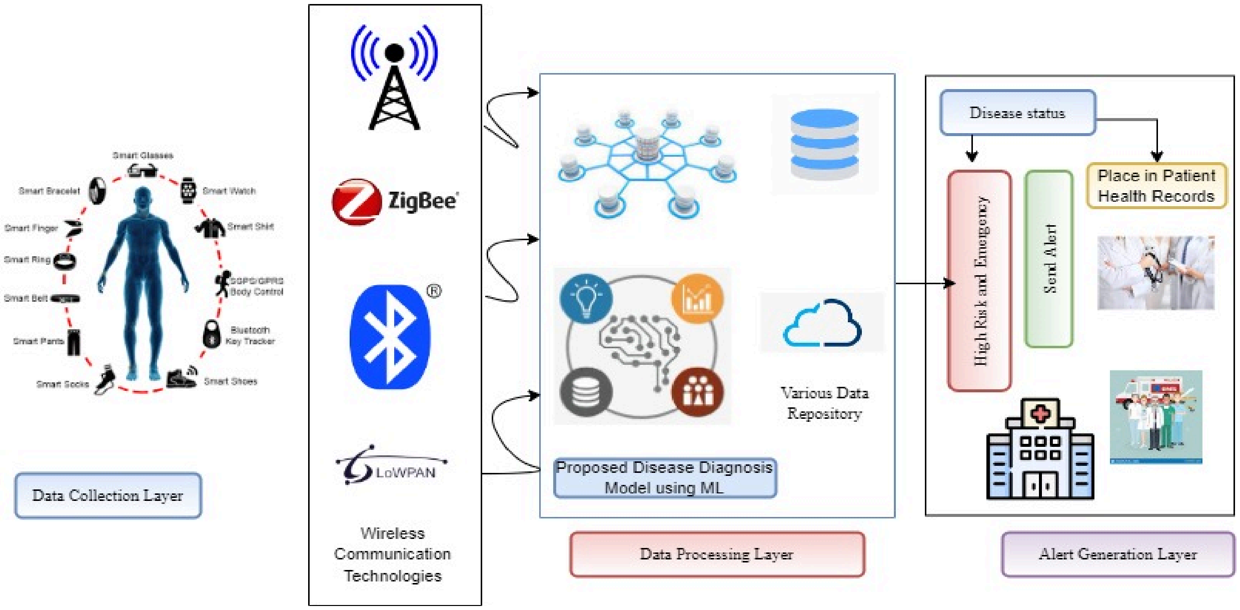


Fig. 2. PD Patients Monitoring and Classification using Wearable Sensors in IoT Environment.

IV. Materials and Methods

This section will describe the data set and techniques utilized for the implementation of our work.

A. PD Dataset Description

Parkinson’s dataset utilized in this study is publicly available at Psyionet [23]. This dataset was collected via a team of three scientists at the “Movement disorder unit of the Tel-Aviv Sourasky Medical Center, Israel” and named after that Yogev et al. [24] consist of the dataset recordings when subjects walking on level ground, Frenkel-Toledo et al. [25] contains dataset recordings when the patient walking on a treadmill and Hausdorff et al. [26] contains recordings when subjects moving at a comfortable place with RAS. This dataset recording was taken from 73 healthy subjects and 93 subjects affected by Parkinson’s disease. In this work, three datasets are separately considered for the identification of healthy and Parkinson's patients.

Table I details the total number of PD and healthy patients from three datasets with their associated physical and clinical characteristics. For small and simple depiction these datasets are denoted as Ga [24], Si [25], and Ju [26]. Each shoe had pressure sensors shown in Fig. 3. Table II depicts the absolute position of sensors in the X-Y coordinate framework. The VGRF signal representation of PD and the non-PD patients is shown in Fig. 4.

The mean and Standard Deviation of each physical feature's age, height, and, weight are taken out to know the average of each characteristic and dispersion of the dataset relative to its mean value.

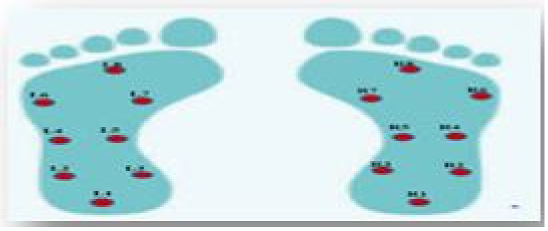


Fig. 3. Pressure Sensors Positioned under each Foot to get the Best Collection.

TABLE I. Number of Subjects in Three Sensors Datasets with their Associated Physical and Clinical Characteristics

Dataset	Group		Subjects		Subject		Avg. (Yrs.) Mean±SD		Height (Mtr.)		Weight (Kg.)	
			Healthy		*PD*		*PD*	*Healthy*	*PD*	*Healthy*	*PD*	*Healthy*
	Healthy	*PD*	*F*	*M*	*F*	*M*						
Ga [24]	18	29	8	10	9	20	61.6±8.8	57.9±6.7	1.67±.07	1.68±.08	73.1 ±11.2	74.2±12.7
Si [25]	29	35	11	18	13	22	67.2±9.1	64.5±6.8	1.66±.07	1.69±.07	70.3±8.4	71.5±11.0
Ju [26]	26	29	14	12	13	16	66.80±10.8	39.31±18.5	1.87±.15	1.83±.08	75.1±11.0	66.8±11.07

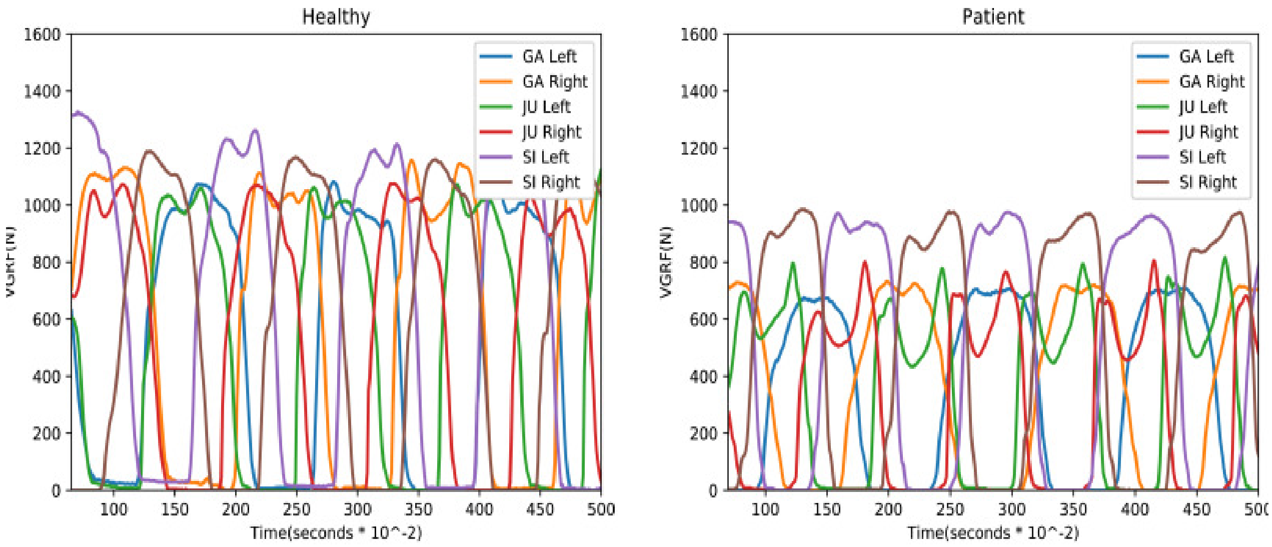


Fig. 4. VGRF Signal Representation of PD Patient and Healthy Person.

TABLE II. PLACEMENT OF LEFT AND RIGHT SENSORS RELATIVE TO X AND Y DIRECTION UNDER EACH FOOT (HERE R DENOTES RIGHT SENSOR, L DENOTES LEFT SENSOR)

Sensor name	**Distance in X- direction (cm)**	**Distance in Y-directionn (cm)**
L1	50	80
L2	70	40
L3	30	40
L4	70	0
L5	30	0
L6	70	40
L7	30	40
L8	50	80
R1	50	80
R2	70	40
R3	30	40
R4	70	0
R5	30	0
R6	70	40
R7	30	40
R8	50	80

B. Proposed ML-based Parkinson's Disease Diagnosis Model

Fig. 5 shows the proposed graphical representation for the diagnosis of PD. First of all, data is collected from wearable eight-foot sensors worn on both left and right feet. After that data may pass through the data pre-processing phase, feature extraction/ selection phase, and final classification phase in which different ML models are applied [31] [32] [33] [34].

- Data Preprocessing

The dataset was collected from different walking tests to avoid the gait starting and ending effect the initial twenty and last twenty data from each gait cycle were removed [35]. The progression of PD among a person can be retrieved through the variations of gait because the walking patterns of individuals changed over time. Therefore, a better study regarding the disease's significant features can provide the best way to understand gait disorders and can be considered significant biomarkers.

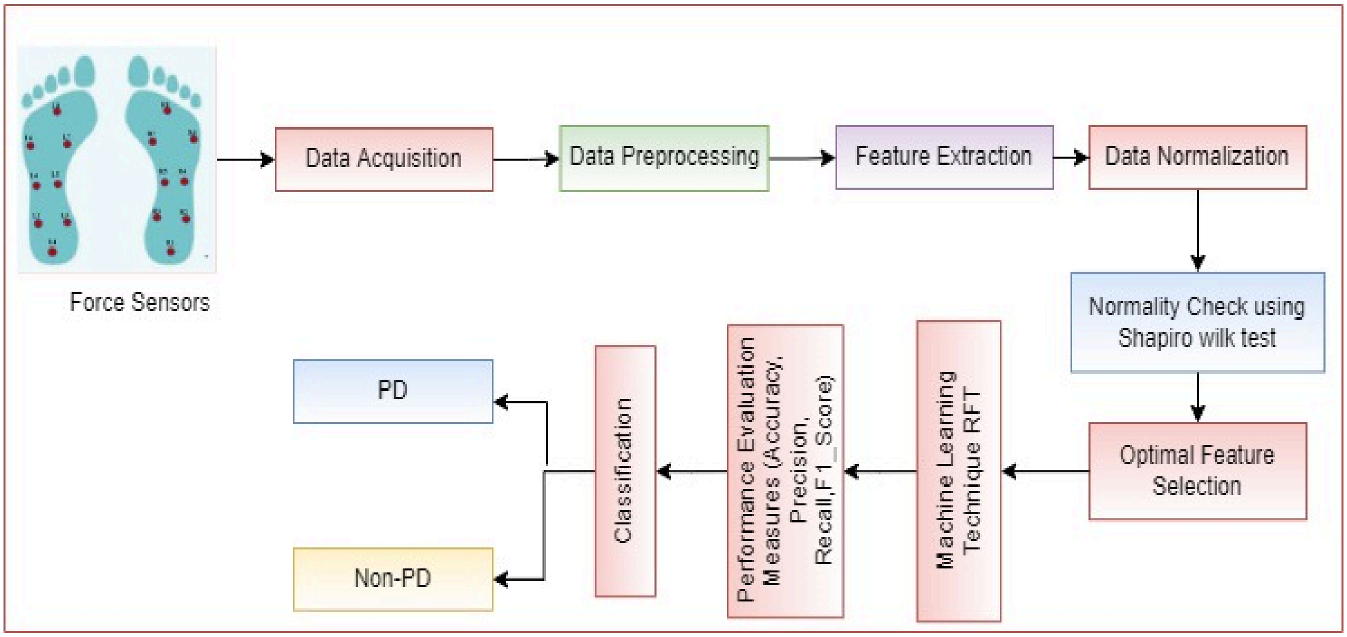

Force Sensors
Data Acquisition
Data Preprocessing
Feature Extraction
Data Normalization
Normality Check using Shapiro wilk test
Optimal Feature Selection
Machine Learning Technique RFT
Performance Evaluation Measures (Accuracy, Precision, Recall,F1_Score)
Classification
PD
Non-PD

Fig. 5. Proposed ML-based Parkinson's Disease Diagnosis Model.

- Feature Extraction

The most important features extracted from raw sensors are depicted in Table III. A Shapiro-Wilk test is utilized to test the normal distribution of features with a confidence bound of 6% for the hypothesis test.

- Data Normalization

The data set was collected from different sensors. Multiple regression techniques are followed to reduce the distribution in the gait data set. That is represented by the equation:

$$Xi = \beta_0 + \sum_{i=1}^{n} \beta_j \, Y_{i,j} + \in i \qquad (1)$$

Where X_i denotes the dependent spatiotemporal features of the ith observation, $X_{i,j}$ denotes the jth physical features like age, weight, height, and walking speed, β specifies the unknown regression coefficient, and $\in i$ denotes the residual error observed for ith iteration. Further feature extraction is performed after normalizing the data.

TABLE III. EXTRACTED SET OF FEATURES WITH THEIR CATEGORY AND NAME

Category	**Feature Name**	**Description**
Time	Stance-duration	The period for which one foot is in direct contact with the floor
	Swing- duration	The period for which the body is completely on the support of one leg
	Stride- duration	The time b/w two continuous events of a similar foot
	Step-time	the time gap b/w starting interaction of one foot to starting contact of contralateral foot
Length	Stride-length	Distance b/w two consecutive ground contacts of the same foot
	Step-length	The gap b/w starting contact of one foot to starting contact of the other foot
Frequency	Cadence	No. of steps occupied per unit of time
Temporal	Swing stance ratio	The proportion of swing to stance interval
	Standardized stance duration (std stn-dur)	The ratio of stance duration to the stride time i.e., (stn-dur/str-dur)
	Standardized swing duration (std-sw-dur)	The ratio b/w swing duration to stride duration, i.e., (sw-dur/str-dur)
	Standardized double limb support (std-DLS-dur)	The ratio among DLS to the stride duration stride time, i.e., (DLS-dur/str-dur)
Force	Heel-strike force	Sensor values mean underneath the heel for initial 5% sample points instance interval of the total gait cycle
	Toe-of force (To-force)	Mean of sensor values underneath the toe for the last 5% sample points instance interval of the total gait cycle
	Centre of pressure (x,y) (COP_x, COP_y)	The total amount of pressure field acting on a body causing a force to act on the ground

- Optimal Feature Selection

Feature selection is a significant step that must be followed before the classification process because it improves overall classification performance and results in less computational time and complexity [30]. A stepwise regression method is applied to select the optimal feature set for classifying the patients into healthy and PD classes. First of all, correlation among various autonomous variables (like gender(G), weight (w), height (h), and walking speed (s)) are calculated by utilizing the Spearman correlation coefficient. Reduction in data dispersion is calculated by utilizing the coefficient of variation with a 95% confidence level (CDL) and a standard error (SE) [31]. The statistical measurable significance of the outcome is evaluated by the value of p as ($p<0.001$). Table IV: describes all selected sets of features.

- Random Forest Tree (RFT)

RFT is introduced by Bierman [31]. RFT is used as a classification technique for this model because analysis of different data mining reveals the RFT as the best one when computed on different datasets in paper [33]. This model works well for both classification and regression-based problems. This method also comes under the ensemble approach as it combines multiple Decision trees. This method was mainly introduced to resolve the pruning problem that occurred in the decision tree approach. Besides searching for the most significant feature while distributing a node, this algorithm looks for the best feature amongst a random set of attributes. RFT method follows the bootstrap aggregating or bagging approach for training the learners. The working procedure of RFT is described in Fig. 6.

TABLE IV. SELECTED SET OF FEATURES WITH NORMALIZED VALUES

Coefficient of Variation (%)	**Raw /Un-normalized Data**			**Standardized Data**		
	ME	***90% CL***	***SE***	***ME***	***90% CL***	***SE***
Cadence	9.28	[8.45:15.47]	1.32	4.98	[4.41:6.55]	0.68
Stride interval	10.45	[8.62:12.26]	0.91	5.41	[5.09:7.71]	0.67
Stride length	17.33	[14.00:20.68]	1.68	5.75	[4.96:6.51]	0.62
Stance interval	13.89	[12.32:15.46]	0.78	6.75	[5.21:7.31]	0.59
Swing interval	11.00	[9.35:12.66]	0.82	9.77	[9.19:11.32]	0.38
Step time	24.21	[21.15:27.33]	1.54	12.29	[11.72:13.81]	0.79
Step length	17.02	[15.67:19.01]	0.83	6.76	[5.98:7.59]	0.58
Double Limb Support	28.70	[27.06: 31.45]	1.41	13.6	[11.07:14.44]	0.35

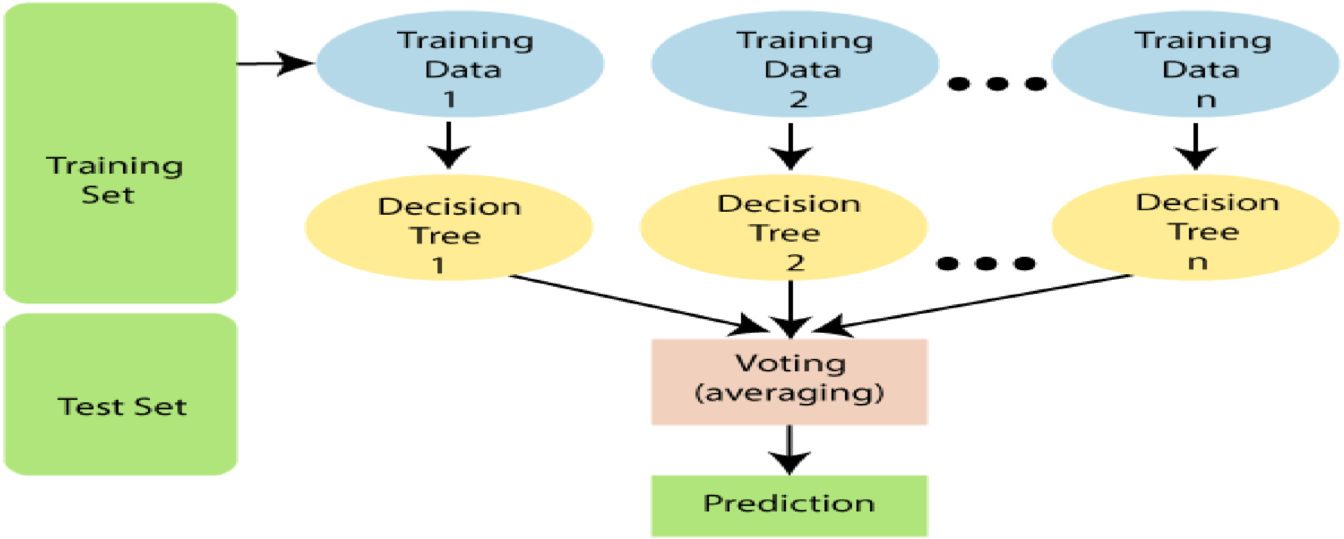


Fig. 6. Working Procedure followed by RFT.

RFT method follow the following steps:

1) For a given training set S=s_1, s_2,……, s_n with classes $Z_{=}z_1,z_{2…,}z_{n,}$ bagging method repetitively (B times) selects a random sample with replacement and applied these samples to fit the tree. For b=1,... B: Selects n training samples from S and Z; and call them S_b, $Z_{b.}$

2) Train the classification tree R_b on S_b, $Z_{b.}$

3) After training phase, predictions for unknown sample S' can be done by taking the majority votes from each classification tree.

4) For a given training set S=s_1, s_2,. ….., s_n with classes $Z_{=}z_1,z_{2…,}z_{n,}$ bagging method repetitively (B times) selects a random sample with replacement and applied these samples to fit the tree. For b=1, ..,B: Selects n training samples from S and Z; and call them S_b, $Z_{b.}$

5) Train the classification tree R_b on S_b, Z_b

6) After training phase, predictions for unknown sample S' can be done by taking the majority votes from each classification tree.

C. Implementation Details

Implementation of the proposed model is done on Jupyter IDE an open-source software developed to support highly interactive data science and scientific computing using Python as a programming language. Most important data computing, visualization, and performance measures and machine learning-based libraries including such as (pandas, NumPy, Matplotlib, sns, metrics, and sklearn) are utilized to support various built-in functionality for computation purposes. Random forest tree (RFT) is utilized as classification techniques to classify subjects into healthy and PD classes and tuned with specific hyperparameters (Max_depth=20, n-estimators=550, criteria=entropy) using Grid search cross validation method [34].

V. Performance Evaluation Metrics

The performances of different ML models can be evaluated by using Accuracy, Recall, Precision, and F1_Score.

- Accuracy

Accuracy mainly refers to the fraction of truly classified samples to the total no. of samples.

$$\text{Accuracy} = \frac{\text{Tn+Tp}}{\text{Tn+Tp+Fn+Fp}} \tag{2}$$

- Recall/Sensitivity

Recall mainly refers to the correctly classified samples by the ML model.

$$\text{Recall} = \frac{Tp}{Tp+Fn} \tag{3}$$

- Precision or Positive Predicted value (PPR)

Precision mainly refers to the proportion of truly classified samples among all positive samples.

$$\text{Precsion} = \frac{Tp}{Tp+Fp} \tag{4}$$

- F1_Score

F1_score ranges between the value 0(refers to worst) and 1(refers to best) and it specifies the balance between recall and precision. It is also called a "weighted harmonic means of precision and recall" and results in an accurate mean of performance of the test.

$$\text{F1\_}Score = 2 \times \frac{\text{Precision}\times\text{Sensitivity}}{\text{Precision+Sensitivity}} \times 100\% \tag{5}$$

VI. Result and Discussion

To avoid the statistically unbiased and overfitting problem, a five-fold CV (Cross-Validation) method is applied. After that, it was arbitrarily divided into five equivalent parts and among five, four subsets are utilized to train the model and the remaining subsets are utilized to test the model. This study addressed a classification problem on three different datasets by utilizing two different feature sets. Table V, VI will show the performance results from three different datasets. The graphical visualization of the performance results is shown in Fig. 7-10. It also shows that there is a great improvement in model performance when is trained with an extracted set of features. For the Ga sub dataset, there is an improvement of approximately 2% can be identified when utilizing feature selection. In the same manner, for the Ju sub dataset, there is an improvement of 4% can be identified when utilizing feature selection technique. Almost the same improvement can be found, for the Si sub dataset, there is an improvement of approximately 2% can be identified when utilizing feature selection. The result revealed that the best combination of related features obtained through feature selection improve the overall performance of proposed ML model. Along with that, it will also reduce execution time both in the training and testing phases.

TABLE V. PERFORMANCE RESULTS BEFORE THE APPLICATION OF THE FS METHOD FOR INDIVIDUAL SUB-DATASET UTILIZING 5-FOLD CV METHOD

Data Sets	**Accuracy**	**Precision**	**Recall**	**F1_Score**
Ga [27]	93.5	88.12	87.4	89.42
Ju [28]	94.42	92.50	90.21	95.60
Si [29]	92.52	90.20	91.21	89.1

TABLE VI. PERFORMANCE AFTER THE APPLICATION OF FS METHOD FOR INDIVIDUAL SUB-DATASET UTILIZING 5-FOLD CV METHOD

Data Sets	**Accuracy**	**Precision**	**Recall**	**F1_Score**
Ga [27]	95.54	91.12	89.4	91.42
Ju [28]	98.80	96.50	95.12	95.24
Si [29]	94.52	91.20	92.12	89.97

It is observed that when proposed model applied on three datasets provide the best accuracy result in the classification of patient into healthy and PD class. Along with that, by evaluating the differences in results between three datasets, it can be notified that proposed ML models show their best results in the case of the Ju dataset. It is also notified that results obtained on Ga are much better as compared to Si. Result can also present the fact that the smaller number of patients with little severity and high severity is more important. It is highly notifying that the patients with little severity may be considered healthy by ML models in some cases and may be considered patients in some cases. High severity patients can be easily separated from healthy subjects.

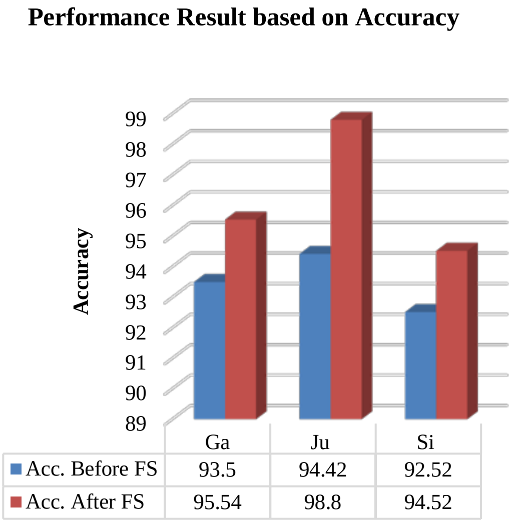


Fig. 7. Accuracy Score with and without Feature Selection.

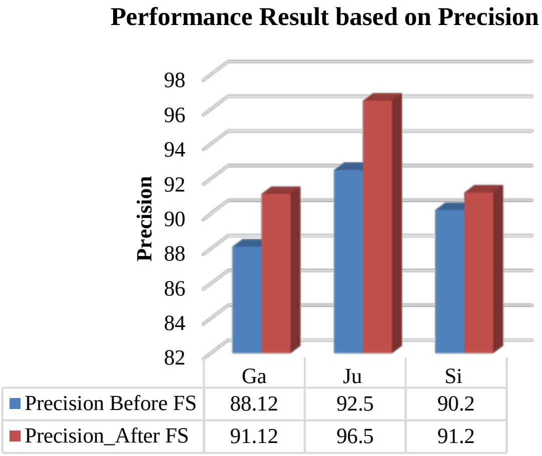


Fig. 8. Precision Score with and without Feature Selection.

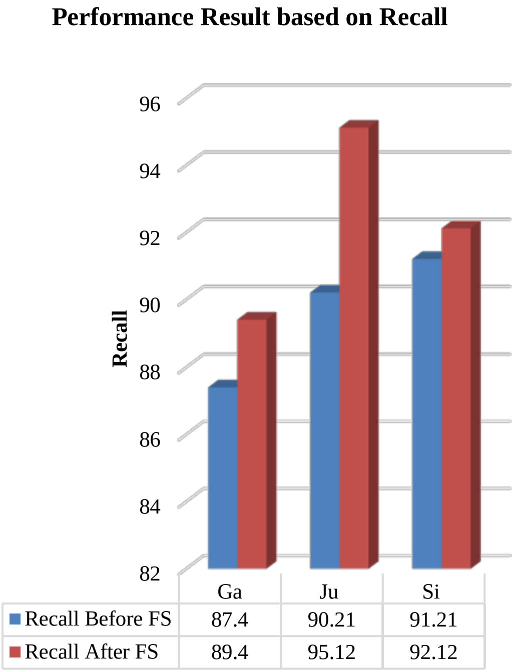


Fig. 9. Recall Score with and without Feature Selection.

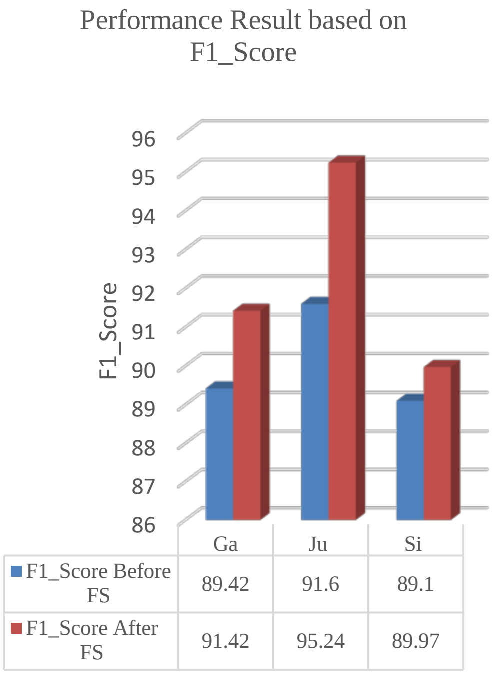


Fig. 10. F1_Score with and without Feature Selection.

Table 7 shows the comparison between the proposed work and the existing work. Results revealed that our proposed work is more precious than the existing work and will support maximum accuracy of 98.80%.

TABLE VII. ACCURACY COMPARISON BETWEEN PROPOSED WORK AND PREVIOUS WORK

Reference	**Features Type**	**No. of Features**	**ML Technique**	**Accuracy Score (%)**
Wu and Krishnan [35]	Time	3	SVM	90.32
Perumal and Sankar [36]	Frequency	10	SVM, NN, LDA	87.52-92.5
Abdul hay et al. [37]	Frequency and Time	3	MEDIUM Gaussian SVM	94.8
Khoury et al. [38]	Spatiotemporal	19	K-NN, DT, RF	83.3-92.86
Alam et al. [39]	Time	33	SVM, K-NN	93.6
Khera et al. [40]	Time Features	10	KNN, SVM, DT, RF	85-98.50
Proposed Work	Spatial, Time and Force	8	RFT	98.80 (Ju), 95.54 (Ga), 94.52 (Si)

VII. CONCLUSION

Wearable sensors offer a smart way of assessing the movement disorders among patients suffering from PD and provide a potentially vast quantity of informative data in order to quantify and monitor the progression of PD among patients. The current work, proposed a ML based PD diagnosis model and evaluated on VGRF dataset (Ga, Si, Ja) composed of different gait cycles that is publicly available on Psyionet. The performance of proposed model assessed by utilizing accuracy, precision, recall and F1_score as performance evaluation measures. The current work, presented that the proposed model outperformed other existing model to classify the patient into healthy and PD classes when subjected to preprocessed gait dataset with selected set of features. A five-fold cross validation method is implemented to achieve the accuracy result of 98.80 on Ju dataset, 95.54 on Ga dataset and 94.52 on Si dataset.

VIII. FUTURE DIRECTION

The conclusion reveals that the proposed model provides best results in the classification of PD but still faces some challenges. Future work concerns the introduction of more relevant features in order to enhance the diagnosis of PD and investigation of the cost-effective hybrid ML/Ensembles ML models and also handling the issue of imbalanced data.

ACKNOWLEDGMENT

I am highly thankful to Dr. Pooja Mittal for her comments and support in the preparation of this manuscript.

REFERENCES

[1] “Statistics|Parkinson’s Foundation” , 2022.

[2] W. Yang, J. L. Hamilton, C. Kopil, C. B. James, M. T. Caroline, L A. Roger, E. R. Dorsey, N. Dahodwala, I. Cintina, P. Hogan, T. Thompson, “Current and projected future economic burden of Parkinson’s disease in the U. S.,” *npj (national publishing group) Parkinson's Dis.* , vol. 6, no. 15 , July 2020.doi: https://doi.org/10.1038/s41531-020-0117-1.

[3] Parkinson’s disease: Causes, Symptoms, Stages, Treatment, Support (clevelandclinic.org), 2022.

[4] A. Z. Khan, F. Aamir, A. Kafeel, M. U. Khan, “Freezing of gait detection in parkinson’s disease from accelerometer readings,” Int. Conf. on Computing, July 2021.

[5] H. Zhonelue, Li. Gen, G. Chao, T. Yuyan, L.Jun, Z. Jin, L.Yun, Yu . Xiaoliu, R. Kang, C. Shengdi, “ Prediction of freezing of gait in parkinson’s disease using a random forest model based on an orthogonal experimental design: a pilot study, “ Front. in Hum. Neurosci., vol. 15, 2021. doi: https://doi.org/10.3389/fnhum.2021.636414.

[6] G. Shalin, S. Pardoel, E. D. Lemaire, J. Nantel, J. Ofman, “Prediction and detection of freezing of gait in Parkinson’s disease from plantar pressure data using long short-term memory neural-networks, “ J. of Neuro Engi. Rehab. , vol. 18, no. 167, 2021. doi: 10.3390/s20154345.

[7] K. M Giannakopoulou, I. Roussaki, K. Demestichas, “Internet of things technologies and machine learning Methods for parkinson’s disease diagnosis, monitoring, and management: a systematic review, “ Sensors, vol. 22, no. 5, 2022. doi: 10.3390/s22051799.

[8] X. Jiang, C. Napier, B. Hannigan, J. J Eng., C. Menon, “Estimating vertical ground reaction force during walking using a single inertial sensor, “ Sensors, vol. 20, no. 15, Aug. 2020. doi: 10.3390/s20154345.

[9] D. Castro, W. Coral, C. Rodriguez; J. Cabra, J. Colorado, “Wearable-Based Human Activity using an IoT approach, ” J. Sens. Actuator Netw., vol. 6, no. 28, 2017. doi: https://doi.org/10.3390/jsan6040028.

[10] E. Abdulhay, N. Arunkumar, and N. Kumaravelu, E. Vellaiappan, “Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson's disease”, Future Gen. Comp. Sys., vol. 83, June 2018. doi: https://doi.org/10.1016/j.future.2018.02.009.

[11] J. M. C. Desrosiers, and J. Frasnelli, “Machine Learning for the diagnosis of Parkinson’s disease: A review of literature,” Front. in aging neurosci., vol. 13, 6 May 2021. https://doi.org/10.3389/fnagi.2021.633752.

[12] P. Drotár, J. Mekyska, I. Rektorová, L., Masarová, Z. Smékal, and M. Faundez Zanuy, “Analysis of in-air movement in handwriting: a novel marker for Parkinson’s disease,” Comp. Methods and Prog. in Biomedicine, vol. 117, no. 3, pp. 405–411, 2014. Doi: https://doi.org/10.1016/j.cmpb.2014.08.007.

[13] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez Zanuy, “Decision support framework for Parkinson’s disease based on novel handwriting markers, “ IEEE Trans. on Neural Syst. and Rehailb. Eng., vol. 23, no. 15, pp. 508–516, May 2015.

[14] P. Drotár J. Mekyska, I, Rektorová, L. Masarová, Z. Smékal, and M. Faundez Zanuy, “Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson’s disease, “ Artif. Intell. in Med, vol. 67, 39–46, 2016.doi : 10.1109/TNSRE.2014.2359997.

[15] H. Choi, S. Ha, H. J. Im, S. H. Paek and D. S. Lee, “Refining diagnosis of Parkinson’s disease with a deep learning-based interpretation of dopamine transporter imaging, “ Neuroimage Clin., vol. 16, pp. 586–594, Sep. 2017. doi : 10.1016/j.nicl.2017.09.010.

[16] L. Silveira-Moriyama, A. Petrie, D. R. Williams, A. Evans, R. Katzenschlager, E. R. Barbosa, and A. J. Lees, “The use of a color-coded probability scale to interpret smell tests in suspected parkinsonism,” Movement Disorders, vol. 24, no. 8, pp. 1144–1153, June 2009. doi: 10.1002/mds.22494.

[17] G. Valenza, S. Orsolini, S. Diciotti, L. Citi, E. P. Scilingo, M. Guerrisi, S. Danti, C. Luchetti, C. Tessa, R. Barbieri, et al., “Assessment of spontaneous cardiovascular oscillations in Parkinson's disease,” Bio. Signal Process. and Control, vol. 26, pp. 80–89, April 2016. Doi: https://doi.org/10.1016/j.bspc.2015.12.001.

[18] F. Wahid, R. K. Begg, C. J. Hass, S. Halgamuge, and D. C. Ackland, ‘‘Classification of Parkinson’s disease gait using spatial-temporal gait features,’’ IEEE J Biomed Health Inform. , vol. 19, no. 6, pp. 1794–1802. doi: 10.1109/JBHI.2015.2450232.

[19] C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. Counsell, & L. Seidl, " Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations the Movement Disorder Society Task Force on rating scales for Parkinson's disease" Mov Disord., vol. 19, no. 9, pp. 1020-1028, Sep. 2004. doi: 10.1002/mds.20213.

[20] L. J. W. Evers, J. H. Krijthe, M. J. Meinders, R. Bloem, T. M. Heskes "Measuring Parkinson's disease over time real-world within-subject reliability of the MDS-UPDRS, " Mov Disord., vol. 34, no. 10, pp. 1480-1487, Oct. 2019.

[21] B. Baker, W.Xiang and I. Atkinson, "Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities, in *IEEE Access*, vol. 5, pp. 26521-26544, 2017, doi: 10.1109/ACCESS.2017.2775180.

[22] M. Raza, M. Awais, S. Hussain, " Intelligent IoT framework for indoor healthcare monitoring of Parkinson's Disease Patient, " in *IEEE Journal on Selected Areas in Comm.* , vol. 39, no. 2, pp. 593-602, Feb. 2021, doi: 10.1109/JSAC.2020.3021571.

[23] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark & H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. vol. 101, no. 23, Jun 2000. doi: 10.1161/01.cir.101.23.e215. PMID: 10851218.

[24] G. Yogev, N. Giladi, C. Peretz, S. Springer, E.S. Simon, J. M Hausdorff, " Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding?", Eur J Neuroscience, vol. 22 , no. 5, pp. 1248-56, Sep. 2005. doi: 10.1111/j.1460-9568.2005.04298.x.

[25] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, J. M. Hausdorff, " Treadmill walking as a pacemaker to improve gait rhythm and stability in Parkinson's disease", Mov Disord., vol. 20(9), pp.1109-1114. Sep. 2005. doi: 10.1002/mds.20507.

[26] J. M Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, N. Giladi," Rhythmic auditory stimulation modulates gait variability in Parkinson's disease", Eur J Neuroscience, vol. 26, pp. 2369-2375, 2007. doi: 10.1111/j.1460-9568.2007.05810.x.

[27] E. Balaji, D. Brindha, V. K. Elumalai and K. Umesh, "Data-Driven Gait Analysis for Diagnosis and Severity Rating of Parkinson's Disease, " Med.l Engi. and Physics, vol. 91, pp. 54-64, May 2021. doi: 10.1016/j.medengphy.2021.03.005.

[28] E. Balaji, D. Brindha and R. Balakrishnan, "Supervised Machine Learning based Gait Classification System or Early Detection and Stage Classification of Parkinson's Disease, " App Soft Compt J., vol. 94, 2020. doi: https://doi.org/10.1016/j.asoc.2020.106494.

[29] R. Prashanth, Sumantra Dutta Roy, Pravat K. Mandal, Shantanu Ghosh, "High-Accuracy detection of early Parkinson's disease through multimodal features and machine learning, " Int J Med Inform., vol. 90, pp. 13-21, 2016. doi: 10.1016/j.ijmedinf.2016.03.001.

[30] J. Sappakitkamjorn, S. A. Niwitpong, "Confidence intervals for the coefficients of variation with bounded parameters", Int J Math and Comput Sci., vol. 7, no. 9, pp. 1416–1421, 2013. doi.org/10.5281/zenodo.1087806.

[31] M. Krzywinski, N. Altman, "Classification and regression trees", Nat Methods, vol. 14, pp. 757–758, Aug. 2017. https://doi.org/10.1038/nmeth.4370.

[32] L. Breiman," Bagging predictors. Machine Learning, " vol. 24, pp.123–140, 1996.

[33] Navita, P. Mittal , "Healthcare Data Analysis using Data Mining Techniques for Disease Prediction", Indian Journal of Comp. Sci. & Eng., vol. 12, no. 5, Oct-Sep 2021.

[34] Scikit Learn. Available online: sklearn.model_selection.Grid Search CV — scikit-learn 1.1.1 documentation.

[35] Y. Wu, S. Krishnan," Statistical analysis of gait rhythm in patients with Parkinson's disease, " IEEE Trans on Neural Syst and Rehabil, vol. 18, no. 2, April 2010. doi: 10.1109/TNSRE.2009.2033062.

[36] S. V. Perumal, R. Sankar,"Gait and tremor assessment for patients with Parkinsons disease using wearable sensors, " ICT Express, vol. 2, no. 4, pp. 168-174, Dec. 2016. https://doi.org/10.1016/j.icte.2016.10.005.

[37] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan, V. Venaatraman, "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease, " FGCS , vol. 83, pp. 366-373, 2018. doi: https://doi.org/10.1016/j.future.2018.02.009.

[38] N. Khoury, F. Attal, Y. Amirat, L. Oukhellou, S. Mohammed, "Data Driven based Approach to aid Parkinson's Disease Diagnosis, " Sensors, vol. 19, no. 2, Jan. 2019. doi: 10.3390/s19020242M.

[39] N. Alam, A. Garg, T.T. K Munia, R. Fazel-Rezai, K. Tavakolian, "Vertical ground reaction force marker for Parkinson's disease, " PloS One, vol. 12, no. 5, May 2017. https://doi.org/10.1371/journal.pone.0175951.

[40] P. Khera, N. Kumar, "Age-Gender Specific Prediction Model for Parkinson's Severity Assessment using Gait Biomarkers", Int.Journal of Engg. Sci. and Tech., vol. 27, March 2022. https://doi.org/10.1016/j.jestch.2021.05.009.

A Novel Methodology for Disease Identification using Metaheuristic Algorithm and Aura Image

Manjula Poojary, Yarramalle Srinivas
Department of Computer Science, GITAM Deemed to be University
Visakhapatnam, India

***Abstract*—Every human has a specific Aura. Every organism in the human body emits energy comprising of Ultra Violet radiation, thermal radiation, and electromagnetic radiation. These energy levels help to underline the physical health inside the human body. In general, these energy levels are called Aura. In order to capture the energy levels, specific cameras like Kirlian are used. These cameras try to capture the energy distribution and map them to the individual organs of the human body. In this article, we present a methodology using Image processing techniques, where Bivariate Gaussian Mixture Model (BGMM) is considered as a classifier to identify the diseases in humans based on the energy distribution. In this article, we have considered five different categories of diseased organs that are identified based on the energy distribution. The preprocessing is subjected to the morphological technique and Particle Swarm Optimization (PSO) algorithm is considered for feature extraction. The segmentation process is carried out using the feature extracted and training is carried out using the BGMM classifier. The result obtained is summarized using various other methods like Support Vector Machine (SVM), Artificial Neural Network (ANN), and Multiclass SVM (MSVM). The results showcase that the proposed methodology exhibits recognition accuracy at 90%.**

Keywords—Aura images; BGMM; image classification; multiclass SVM; artificial neural network

I. Introduction

Image processing is the primary step in image analysis. It helps to both process the image, and enhance the image and also helps in the effective recognition of deformities inside the image regions based on the feature extraction techniques [1]. In image processing the primary objective is to process the input image so that effective recognition can be achieved [2]. Among the various application of image processing recently much emphasis is subjected to the area of medical imaging. In medical imaging the acquired medical images are processed, enhanced and the features are extracted for the effective identification of the disease [3, 4]. Of late, many article have been proposed for effective identification of image deformities based on the medical images as inputs. Some of the techniques in this area of research include; methodologies based on non-parametric technique such as pixel-based techniques, region-based techniques, shape based, texture-based techniques etc. [5, 6, 7, 8]. The research is also extended further by developing parametric models for better classification and identification of medical diseases using methodologies like GMM, Hidden Markovian Models, Markovian Random fields etc. [9, 10, 11, 12]. It is also presented in the literature that parametric model-based approach is more adaptive compared to the non-parametric model [13, 14, 15, 16]. The latest advancement in the area of computer communication technologies led towards the development of machine learning technique and thereby literature has been driven in this direction of research, using machine learning technique such as Convolution Neural Network (CNN), Artificial Neural Network (ANN), Bayesian belief network etc. [17, 18, 19, 20]. However, in spite of the numerous works presented by the authors, effective identification and diagnosis is only subjected after the acquisition of medical scans and by then the disease might have entered the human anatomy. However very little work is reported in the literature to estimate the deformities inside the human body well before the disease is noticed.

This article attempts in this area of research where the Aura images are considered and using the Aura images as the input, prior detection of the disease can be identified even before the development of the disease. For this purpose, we have considered the database of Aura images, namely; Bio-Well data set consisting of Aura images. In this article we have tried to highlight the diseases like nervous system, thorax, thyroid, abdomen disease and throat diseases. The article is further presented in the following sections, where Section II deals with the brief introduction to Aura images, Section III deals with dataset considered. Feature extraction based on the PSO is highlighted in Section IV of the article. Section V deals with classification technique based on BGMM and methodology is highlighted in the corresponding Section VI. The results derived are compared with the existing algorithm and are presented in corresponding results Section VII of the article. Section VIII of the article summarizes with conclusion.

II. Aura Images

The Aura images are nothing but the energy field surrounding the human. Every individual is attributed with a specific Aura. The Aura images exhibit different colors in tandem with the colors of VIBGYOR, where each color is related to the chakras of the human body. The violet color in the colors of VIBGYOR indicates crown chakra, indigo specifies the third eye chakra, blue indicates throat chakra, green indicates the heart chakra, yellow color indicates the solar plexus, orange color indicates the sacral chakra and red indicates the base chakra. Each of the colors corresponding to the chakras are linked with an organ in the human body and they exhibit the behavior of the individual. The red color indicates the anxiety levels and the orange color represents the structure of the kidney or reproductive organ, yellow is subjected to the spleen, the working condition of lungs can be identified by the green color, throat and thyroid are reflected by the blue color, the nervous system problem can be identified by the violet and indigo color is subjected to the problem related to lungs. The main advantage of the Aura images is that, if the intensity around each chakra is identified, the specific disease which is likely to appear can be known in prior by the higher intensity depicted in these colors [20, 21].

III. Bio-Well Dataset

Bio-Well dataset records a large number of human aura images that are captured using a device called Gas Discharge Visualization (GDV), which is based entirely on Electro photonic imaging (EPI). These images are used for imaging analysis to identify many of the physical problems associated with the person. The GDV device captures the electronic cloud on the finger, those electronic images or finger images are transmitted to the GDV device for image processing. The images obtained from GDV are processed. The Bio-Well database includes aura images associated with the fingers, full aura images of the body plus aura images associated with the chakras of the human body. Bio-Well image provides a person with a wide range of seals and observations about strength and endurance for the whole body, energy centers, organs and systems. Bio-Well is used by hundreds of physicians, clinicians and researchers around the world.

IV. Particle Swarm Optimization Algorithm (PSO)

PSO is metaheuristic, populace is primarily based totally on stochastic seek set of rules stimulated with the aid of using social conduct of birds flocking at the same time while trying to find the food. Here every particle represents bird and they are able to fly in unique route and every particle has speed and position and swarm represents group of birds or populace. The idea of food searching behavior of birds is used to mathematically model the algorithm. This algorithm is used to solve the optimization problem and PSO seeks for maximum cost with the aid of using the iterations.

Suppose group of birds are randomly flying and attempting to find the food in a place (search area), social behavior of the birds is moving in the direction of a crowd. Assume that there is only one piece of food in a place being searched and all of the birds do not know the position of the food however they know how far they're in each iteration.

The search procedure used are

- Follow the bird that's nearest to the food.
- Birds do not know the best position.
- If any member can locate the appropriate path, rest of the members will comply with quickly. Starting with the randomly initialized populace and moving in randomly initialized directions, every particle is going through the search area and remembers the best preceding positions of itself and its neighbors. That is every particle seeks for maximum value through updating the iterations. In every iteration each particle is updated following the two best values, .i.e. pbest (best particle position) and gbest (group best particle position).

A. Mathematical Model

In PSO every particle is taken into consideration to be solution and every particle has its speed, position and fitness values. In swarm every particle remembers its position called particle_bestFitness_value(pbest), particle_bestFitness_position. A record of global_bestFitness_position(gbest) and global _bestFitness_value is maintained. Position and velocity is calculated using (1) and (2)

$$\mathrm{x}_\mathrm{i}^{\mathrm{t+1}} = \mathrm{x}_\mathrm{i}^\mathrm{t} + \mathrm{v}_\mathrm{i}^\mathrm{t} \tag{1}$$

in which x_i^t is the previous position of the particle, v_i^t *is* particle velocity and x_i^{t+1} is current position.

$$\mathrm{v}_{\mathrm{k=1}}^\mathrm{t} = \mathrm{wv}_\mathrm{k}^\mathrm{i} + \mathrm{c}_1\mathrm{r}_1(\mathrm{xBest}_\mathrm{i}^\mathrm{t} - \mathrm{x}_\mathrm{i}^\mathrm{t}) + \mathrm{c}_2\mathrm{r}_2(\mathrm{gBest}_\mathrm{i}^\mathrm{t} - \mathrm{x}_\mathrm{i}^\mathrm{t}) \tag{2}$$

here w is inertia weight, c1, c2 is positive constants, r1,r2 is random values within the range[0,1].

$xBest_i^t$ is best particle position and $gBest_i^t$ is global best.

Calculate the fitness value of every particle using the objective function and pick out the best fitness value as gbest.

Algorithm

1. Initialize parameter and population
2. Calculate fitness value (optimum) for every particle. If the fitness value is the higher than best fitness value (pbest) then set new value as global best (gbest). Choose the particle with high-quality fitness value as gbest.
3. For every particle calculate speed and position. Calculate fitness value and locate gbest.
4. Repeat this method till circumstance met.
5. Replace counter t=t+1
6. Output is gbest(global best) and x_i(position)

V. Bivariate Gaussian Mixture Model (BGMM)

The image segmentation is the primary factor for image analysis and retrieval. Various colors are proposed for different image processing scenario. For various image processing scenarios, different colour models are suggested. Through the use of hue and saturation in a bivariate Gaussian mixture model (BGMM), human perception of a picture can be described in HIS (hue saturation intensity) colour space. The image is regarded as a finite mixture of BGMM for the purpose of image segmentation, with the feature vector of each image region having a bivariate Gaussian distribution. Through the use of a bivariate frequency surface and the K-means algorithm, the number of components in the mixture are determined. The model parameters are calculated by deriving the revised equations of the model parameters for the EM-algorithm. The segmentation is carried through maximizing the component likelihood. It is common to assume that the feature vector of the image follows a bivariate Normal (Gaussian) distribution when modelling the bivariate features of the image. The Probability density function of the BGMM is given by.

$$\mathrm{f}(\mathrm{x}_1,\mathrm{x}_2) = \frac{1}{2\pi\sigma_1\sigma_2(\sqrt{1-\mathrm{p}^2})} e^{-\left[\frac{1}{2(1-\rho^2)}\left[\left(\frac{\mathrm{x}_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\mathrm{x}_1-\mu_1}{\sigma_1}\right)\left(\frac{\mathrm{x}_2-\mu_2}{\sigma_2}\right) + \left(\frac{\mathrm{x}_2-\mu_2}{\sigma_2}\right)^2\right]\right]} \tag{3}$$

where μ_1, μ_2 are any real numbers.

$\sigma 1 > 0$, $\sigma 2 > 0$; $-1 <= \rho <= 1$

in which μ1, σ1 are the mean and variance of the image with 1st features and μ2, σ2 are the mean and variance of the image with the 2d features, ρ is referred to as the shape parameter.

If $\rho = 0$ this implies correlation (x, y) = 0. i.e.

$$\mathrm{f(x1, x2)} = \frac{1}{2\pi\sigma 1\sigma 2(\sqrt{1-p2})} \mathrm{e}^{-\left[\frac{\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}{2}\right]} \tag{4}$$

$$= \mathrm{f}(x_1)(x_2) \tag{5}$$

x and y are normal variables.

Equations for first feature is

$$\sigma_1 = \frac{K_1 \pm \sqrt{{K_1}^2 - 4K_2}}{2} \tag{6}$$

$$where\ K_1 = \left(\frac{2e(x_1-\mu_1)(x_2-\mu_2)}{\sigma_2}\right)\left(\frac{-1}{2(1-\rho^2)}\right) \tag{7}$$

$$and\ K_{2=} \frac{2\,(x_1-\mu_1)^2}{(1-\rho^2)} \tag{8}$$

Equation for second feature is

$$\sigma_2 = \frac{-K_3 \pm \sqrt{{K_3}^2 - 4K_4}}{2} \tag{9}$$

$$\text{where } K_3 = \frac{1}{2(1-\rho^2)} \frac{2e\,(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1} \tag{10}$$

$$and\ K_4 = \frac{-\,(x_2-\mu_2)^2}{(1-\rho^2)}$$

VI. Methodology

The analysis of the disease based on the Aura images is done using the BGMM classifier by adopting the following steps.

Step 1: Morphological processing

The first step in the identification of the disease is to acquire the image and process the image. In order to have a precise enhancement it is customary to process the image by the integration of morphological technique together with segmentation so as to acquire the inherent outer Auras effectively [22]. The input image is converted to binary image and using the technique of morphology such as erosion and dilation the input image is processed. This process enhances the quality of the image and in order to further enhance the quality of the image technique like closing, Hit and Miss are used for effective smoothing and filling the holes inside the images. These processed images are considered for segmentation.

Step 2: Segmentation

In this process the images are clustered into groups such that the homogeneous pixels are together in a specific group based on a criterion obtained using the color intensity. These intensities of pixels are given as input to the BGMM so that high intensity regions can be identified. These regions with high intensity attribute to the deformity.

Step 3: Feature extraction

The methodology carried out in two phases training and testing which is shown in Fig. 1. Each aura image of individual extracted and the high intensity levels identified from the BGMM are used to recognize the disease.

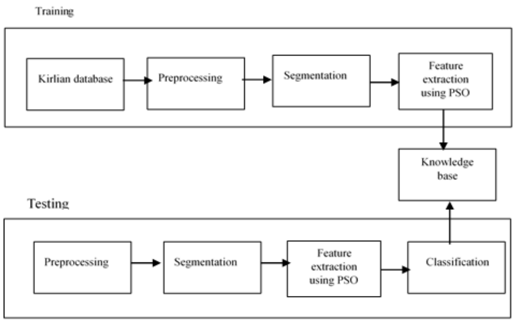


Fig. 1. Model of the Proposed Methodology.

VII. Result and Discussion

In order to present the results, the Aura images are classified based on the intensity levels. Different techniques such as ANN, MSVM and PSO with BGMM are considered for undergoing the comparative study and the various images considered are presented in Fig. 2-11. Fig. 2 shows the depleted energy distribution around neck, abdomen, leg and back. Fig. 3 indicates the energy holes in various regions due to pressure. Fig. 4 shows health issues in head, heart and abdomen region. Disease regions are indicated with breakage in the images region indicates the health problems in throat, heart, stomach and leg area due to hyper tension is shown in the Fig. 5. The aura image which was taken before meditation is given in Fig. 6. The different aura images based on the chakras is given in Fig. 7 to 11. The Fig. 7 shows the high energy levels in the heart and head regions. The problem in digestive system is observed in the Fig. 8. Health issues in the various organs shown in the Fig. 9. The Fig. 10 shows imbalanced energy level in root, solar plexus and throat chakras. Fig. 11 shows the high energy level in the root, sacral and throat chakras indicates health issues in related organs.

All the aura images are obtained from biowell data set. These images are captured using Kirlian based device called Gas discharge visualization (GDV).

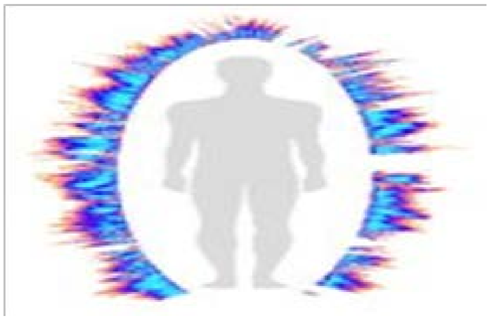


Fig. 2. Energy Depletion in Leg, Abdomen, Back and Head.

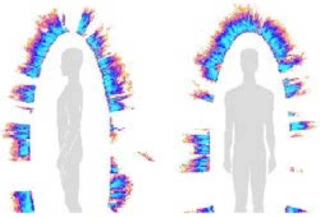


Fig. 3. Energy Holes in Various Organs.

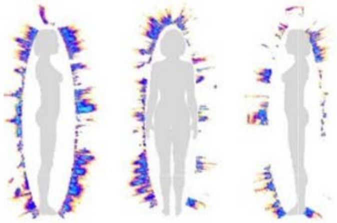


Fig. 4. Energy Depletion in the Region of Head, Heart and Abdomen.

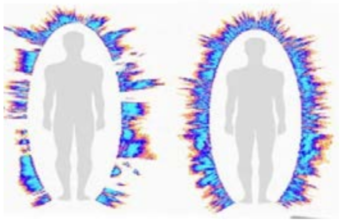


Fig. 5. Energy Holes in the Region of Throat, Heart, Stomach and Leg Area.

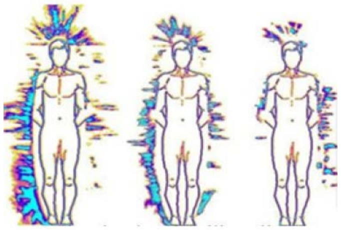


Fig. 6. Diseased Aura Image in Various Organs.

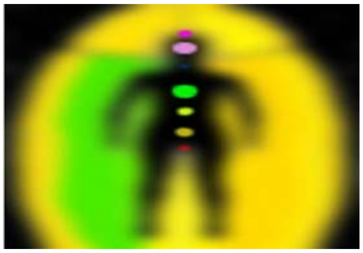


Fig. 7. Health Issues in Heart and Head (Migraine).

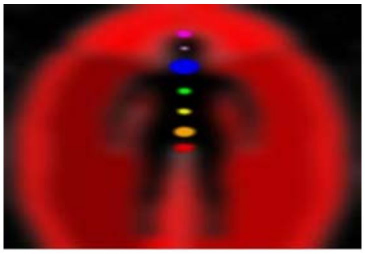


Fig. 8. Health Issues in Throat and Digestive System.

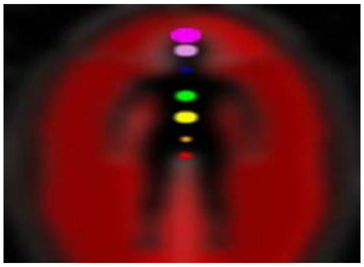


Fig. 9. Health Issues in Pineal Gland, Nervous System, Heart and Kidney.

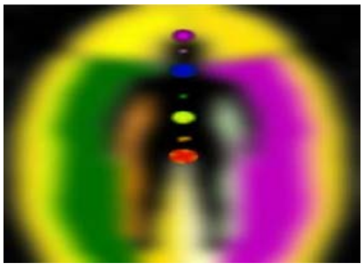


Fig. 10. Health Issues in Addrenal Gland, Solar Plexus , Throat and Brain.

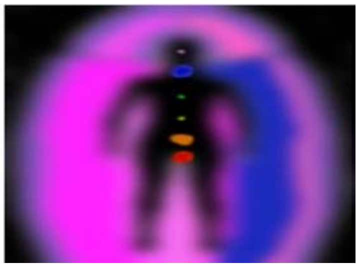


Fig. 11. Problems in Throat, Urogenital System and Spinal Cord.

The comparative analysis is carried out using technique based on ANN, MSVM technique and the proposed model based on PSO. The organs considered for the identification of the disease are nervous system, thorax, thyroid, abdomen and throat. The results obtained are depicted in the Table I. The result shows that proposed method has greater accuracy when compared to the existing model.

TABLE I. COMPARATIVE STUDY OF PROPOSED METHOD WITH OTHER METHODS

Organs having Problem	**No. of samples Trained**	**No. of samples tested**	**No. of samples recognized with ANN**	**No. of samples recognized with MSVM**	**No. of samples recognized with Proposed model**	**Accuracy (%)**
Nervous system	5	43	24	29	39	90%
Thorax	6	55	30	39	47	86%
Thyroid	7	34	12	19	29	87%
Abdomen	5	65	34	39	59	91%
Throat	7	39	21	29	32	83%

From the table it is observed that the proper methodology based on PSO helps to identify the diseases in advance compared to the other methods.

VIII. Conclusion

Aura images have a capability of disease identification. These images can be considered for the identification of the disease with respect to a specific organ. Both testing and training phases are carried out where the test image is identified based on the intensity whether it is prone to disease or not. A new methodology is considered for feature extraction based on BGMM together with PSO. The Bivariate feature H and S are given as input to the model, derived primarily based on Bivariate Gaussian distribution. It helps to identify the color intensities and thereby helps in identifying the high intensity levels resembling the disease. The comparative analysis carried out shows the proposed method helps in effective recognition of the diseases well in advance when compared to other state-of-art methods.

References

[1] Manjula Poojary and Srinivas Yarramalle, “Review of image analysis based on aura images," International Journal of Science ,Engineering and Technology, vol. 8, no. 4, pp. 1-11, 2020.

[2] Xanadu C. Halkias and Petros Maragos, “Analysis of kirlian images: feature extraction and segmentation," Proceedings of ICSP, 2004.

[3] Roeland Van Wijk and Utrecht Eduard P.A, “An introduction to human biophoton emission,” Forsch Komplementarmed Klass Naturheilkd, Vol.12(2), pp. 77-83, April 2005, DOI: 10.1159/000083763.

[4] Loo Chu Kiong and Tehjoo Peng, “Quantum bioinspired invariant object recognition model on system-on-a-chip (soc), IEEE Conference on Robotics, Automation and Mechatronics, pp. 433-438, 2008, doi: 10.1109/RAMECH.2008.468138.

[5] Sitizura A. Jalil, Mohd Nasir Taib, Hasnain Abdullah, and Megawati, “Frequency radiation characteristic around the human body,” IJSST, vol. 12, no. 1, DOI:10.5013/IJSSST.a.12.01.05.

[6] B. Shanmugapriya and R. Rajesh, “Understanding abnormal energy levels in aura images,” Icgst Aiml-11 Conference, Dubai, Uae, pp. 12-14 April 2011.

[7] Janifalalipal, Razakmohd Ali Lee, and Alifarzamnina, “Preliminary study of kirlian image in digital electrophotonic imaging and its application,” 2nd International Conference On Automatic Control And Intelligent System, Kota, Kinabalu, Malaysia, pp. 213-217, 2017, DOI: 10.1109/I2CACIS.2017.8239060.

[8] A.F Abijanska and D. S Ankowski, “Aura removal algorithm for high-temperature image quantitative analysis systems,” 14th International Conference on Mixed Design of Integrated Circuits and Systems, pp. 617-621, June 2007, doi: 10.1109/MIXDES.2007.4286236.

[9] Masaki Kobayashi, Daisuke Kikuchi, and Hitoshi Okamura, “Imaging of ultraweak spontaneous photon emission from human body displaying diurnal rhythm,” PLoS One, vol. 4, no. 7, July 2009, DOI:10.1371/journal.pone.0006256.

[10] Chao-Hui Huang and Danielracoceanu, “Bio-Inspired computer visual system using gpu and visual pattern assessment language (vipal): application on breast cancer prognosis,” The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1-8, 2010, DOI:10.1109/IJCNN.2010.5596972.

[11] K. G. Korotkov, Peter Matravers, Dmitry V. Orlov, and M.S. Bernard O. Williams, “Application of electrophoton capture (epc) analysis based on gas discharge visualization (gdv) technique in medicine: a systematic review,” The Journal of Alternative And Complementary Medicine, vol. 16, no. 1, pp. 13–25, 2010, DOI: 10.1089/acm.2008.0285.

[12] Vinitha sree subbhuraam, E. Y. K. Ng, G. Kaw, and Rajendraacharya, “Evaluation of the efficiency of biofield diagnostic system in breast cancer detection using clinical study results and classifiers,” J Med Syst vol. 36, pp. 15–24, 2012, Doi 10.1007/S10916-010-9441-Z.

[13] R. Rajesh, B. Shanmugapriya, J. Satheesh Kumar, and V.Arulmozhi, “Could aura images can be treated as medical images?,” Internation conference on informatic enginnering and information science, Icieis,vol. 252, pp. 159–170, 2011, https://doi.org/10.1007/978-3-642-25453-6_15.

[14] Nataliya Kostyuk, Phyadragren Cole, Natarajan meghanathan, Raphael D. Isokpehi, and Hari H. P. Cohly, “Gas discharge visualization: an imaging and modeling tool for medical biometrics”, International Journal Of Biomedical Imaging, vol. 2011, Article Id 196460, 2011, Doi:10.1155/2011/196460.

[15] K. Priyadarshini, P. Thangam, and S. Gunasekaran, “Kirlan images in medical diagnosis: a survey,” Int. Journal Of Computer Applications, Proceedings of International conference on Simulation in Computing Nexus, pp. 5-7,2014.

[16] John A. Ives et al., “Ultraweak photon emission as a non-invasive health assessment: a systematic review,” PLoS One, vol. 9, no. 2, 2014, DOI: 10.1371/journal.pone.0087401.

[17] Shreya Prakash, Anindita Roy Chowdhury, and Anshu Gupta, “Monitoring the human health by measuring the biofield "aura": an overview,” International Journal of Applied Engineering Research, Issn 0973-4562, vol. 10, no. 35, pp. 27654-27657 2015, https://www.researchgate.net/publication/277575681.

[18] G. Chhabra, N. Aparna, S. Souvik, “Human aura: a new vedic approach in IT,” International Conference On Mechanical And Industrial Engineering, New Delhi, May 2013, https://www.researchgate.net/publication/280876417.

[19] Erminia Guarneri and Rauni prittinen king, “Challenges and opportunities faced by biofield practitioners in global health and medicine: a white paper,” Globaladv Health Med, vol. 4, pp. 89-96, 2015, Doi: 10.7453/Gahmj.2015.024.Suppl.

[20] Himanshu kharadi and Kritika jain, “Aura-Bio energy,” Proceedings Of International Conference On Emerging Technologies In Engineering, Biomedical, Management And Science, March 2016.

[21] Zhuo Wang, Niting Wang, Zehua Li, Fangyan Xiao, and Jiapei Dai, “Human high intelligence is involved in spectral redshift of biophotonic activities in the brain,” Proc Natl Acad Sci U S A, vol. 113, no.31,pp. 8753-8, Aug 2016, doi: 10.1073/pnas.1604855113.

[22] Rai sachindra Prasad, Shishir Prasad, and Vikasprasad, “Pattern recognition in thought-form images using radon transform and histograms,” Proceedings of the 2nd International Conference on Biomedical Signal and Image Processing(ICBIP), pp. 22-28, 2017, https://doi.org/10.1145/3133793.3133806.

Enhanced Gradient Boosting Machines Fusion based on the Pattern of Majority Voting for Automatic Epilepsy Detection

Dwi Sunaryono[1], Riyanarto Sarno[2], Joko Siswantoro[3]*, Diana Purwitasari[4], Shoffi Izza Sabilla[5], Rahadian Indarto Susilo[6], Adam Abelard Garibaldi[7]
Department of Informatics Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia[1, 2, 4, 7]
Department of Informatics Engineering, University of Surabaya, Surabaya, Indonesia[3]
Department of Medical Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia[5]
Department of Neurosurgery, Dr. Soetomo Academic General Hospital, Surabaya, Indonesia[6]

***Abstract*—Automatic detection of epilepsy based on EEG signals is one of the interesting fields to be developed in medicine to provide an alternative method for detecting epilepsy. High accuracy values are very important for accurate diagnosis in detecting epilepsy and avoid errors in diagnosing patients. Therefore, this study proposes the Enhanced Gradient Boosting Machines Fusion (Enhanced GBM Fusion) for automatically detecting epilepsy based on electroencephalographic (EEG) signals. Enhanced part of GBM Fusion is the pattern of majority voting evaluation based on the fusion of five-class and two-class GBM, called Enhanced GBM Fusion. The raw signal is extracted using Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT), then feature is selected by using Genetic Algorithm (GA) before classification. This proposed method was evaluated using five classes (normal in open eyes, normal in close eyes, interictal with hippocampal, interictal, and ictal) from the University of Bonn. The experimental results show that the proposed Enhanced GBM Fusion can increase the accuracy of GBM Fusion of 99.8% to classify five classes of epilepsy based on EEG signal. However, the performance of Enhanced GBM Fusion cannot be generalized to other datasets.**

Keywords—Epilepsy; enhanced gradient boosting machine fusion; electroencephalographic (EEG) signal; discrete wavelet transform (DWT); discrete fourier tansform (DFT); genetic algorithm (GA)

I. INTRODUCTION

Globally, WHO estimates that about five million people are diagnosed with epilepsy yearly. Low and middle-income countries are nearly three times more than high-income countries to be diagnosed with epilepsy [1]. This is feasible due to the increased risk of endemic conditions, variations in medical infrastructure, the availability of preventive public health programs, and accessible healthcare services.

Epilepsy is a chronic medical disorder with clinical symptoms and signs due to intermittent brain function disorders. It occurs due to abnormal or excessive electrical discharge from neuron paroxysms of various etiologies. Generally, epilepsy is attended in unpredictable, unprovoked recurrent seizures that affect a variety of mental and physical functions. Seizure is a spontaneous electrical hyperactivity activity of a group of nerve cells in the brain and is not caused by an acute brain disease. “Seizure” in epilepsy is an incurable disease. However, about 70% of people with epilepsy can be free from seizures with proper treatment. Neuroscientists generally predict seizures based on an abnormal electroencephalographic (EEG) pattern in the brain. An EEG is a device that records activity in the brain, including seizures.

A method with high accuracy is needed to detect whether the ongoing seizure is an epileptic seizure or a fake seizure. Visual examination of the EEG signal is necessary to determine the occurrence of epilepsy. Unfortunately, checking the EEG signal manually takes a long time, and sometimes the results are missed or false-alarm detections [2]. Automatic epilepsy detection has been studied since the 1970s in the form of literature in the hope of helping the medical world in detecting automatic epilepsy based on EEG data [3]. In general, automatic epilepsy detection can be categorized into two groups which are conventional approaches and Deep learning approaches.

Misdiagnosis of epilepsy is a fatal error because it can lead to inappropriate treatment and death. Therefore, a high accuracy value is essential in the automatic detection of epilepsy. The automatic detection of epilepsy from five classes of EEG data has been tried using various approaches, but not many of them achieve accuracy above 98%. This study presented a method for automatic epilepsy detection based on the motivation that the method can classify five classes of EEG signals with greater than 98 percent accuracy. This study utilized a combination of Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT) for feature extraction in the frequency and time-frequency domains. The output from the feature extraction by DFT and DWT is extracted again using a statistical feature and crossing frequency features. To obtain the best features that would be used for classification, feature selection based on a Genetic Algorithm (GA) is used in this study. This study proposes a method to enhance GBM Fusion that was previously used by Sunaryono et al. [4].

The distinguish of this study from the study conducted by Sunaryono et al. [4] is that this study proposes the Enhanced Gradient Boosting Machine Fusion, where the method will focus more on errors that occur in the classification results. Errors in this classification will be evaluated to see if there is a

*Corresponding Author

pattern in the errors. The pattern will be applied to the Enhanced Gradient Boosting Machine Fusion classification for higher accuracy. The workflow for the proposed method can be seen in Fig. 1. By using the proposed method, this study hopes to have good accuracy results (above 98%) and the study also hopes to contribute to helping the medical world to detect epilepsy automatically so that people with epilepsy can be diagnosed accurately and quickly.

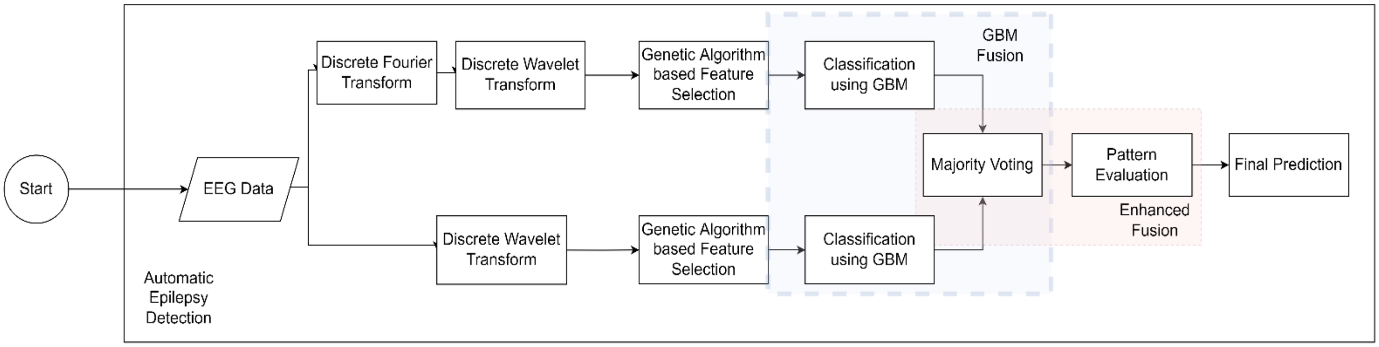

Start
EEG Data
Discrete Fourier Transform
Discrete Wavelet Transform
Genetic Algorithm based Feature Selection
Classification using GBM
GBM Fusion
Discrete Wavelet Transform
Genetic Algorithm based Feature Selection
Classification using GBM
Majority Voting
Pattern Evaluation
Final Prediction
Enhanced Fusion
Automatic Epilepsy Detection

Fig. 1. Flowchart of the Proposed Method.

The remaining sections are organized as follows. Section II provides related works. The Section III outlines the materials and methods utilized in this study. In Section IV, the experimental result and discussion are presented. In Section V, the conclusion is reached.

II. Related Work

This section provides literature study based on the University of Bonn dataset: Wu et al. [5] carried out their research using the Complete Ensemble Empirical Method (CEEMD) to decompose signal data into 12 IMFs and one residue. XGBoost was used as a classification method where the results detected an accuracy higher or equivalent to 99% in 12 cases of two and three classes. Ullah et al. [6] use a Pyramidal One-Dimensional Deep Convolutional Neural Network (P-1D-CNN) as an architecture to perform feature extraction to classify three classes. To improve accuracy, the author added majority voting on the architecture, which has significantly increased the accuracy of P-1D-CNN. This architecture yields an accuracy of 99.1 ± 0.9% in the two and three-class cases.

Türk and Özerdem [7] performed Continuous Wavelet Transform (CWT) as a feature extraction method on EEG signals to get a time-frequency 2-D scalogram image. The scalogram images were used as input for CNN classification. The classification result with the highest accuracy was 99% for three classes and 93.6% for five classes. Wang et al. [8] propose the Discrete Wavelet Transform (DWT) method as a feature extraction method and uses the Gradient Boosting Machine and Grid Search Optimization to optimize the hyperparameters. The study resulted in an accuracy of 96.5% for classifying three classes using a dataset from the University of Bonn.

Sunaryono et al. [4] also suggested using Discrete Wavelet Transform (DWT) in automatic epilepsy detection. Gradient Boosting Machine fusion (GBM Fusion) method is proposed to increase the accuracy of classifying three classes to 99.45%. The proposed fusion combines the results of the classification of 2 classes and three classes using majority voting.

Singh and Dehuri [9] produced 100% accuracy in two and three classes case. Also, they produced 93.33% in five classes case with a hybrid technique using DWT-based Singular Value Decomposition fuzzy *k*-nearest neighbor (SVD-F*k*NN) classifier. On a substantial scale, DWT-based SVD decomposes the input EEG signals into sub-bands. The extracted feature is classified using several '*k*' values for the F*k*NN classifier.

Zhao et al. [10] implemented a 1-D DNN based on CNN for robust automatic epilepsy detection that consisted of three convolutional blocks and three fully connected layers. In addition, each convolutional block consists of five distinct layer types. The proposed method achieved an accuracy of 93.55% in the five classes problem. Zhao et al. [11] continue the research with proposed a method called SeizureNet based on CNN that utilizes two convolutional neural networks to extract features and a fully connected layer to learn high-level features. This method has achieved 95.84% accuracy in five classes case.

Sukriti et al. [12] used two entropy features, called refined composite multi-scale dispersion entropy (RCMDE) and multi-scale dispersion entropy (MDE) to detect seizure from EEG data with one way analysis of variance (ANOVA) as a feature selection method before being classified by Support Vector Machine (SVM). The best accuracy achieved is 96.67% in three classes case using RCMDE.

Zhang et al. [13] modeled a multi-scale non-local (MNL) network, 1 D CNN, to identify epilepsy automatically. Signal pooling and multi-scale non-local layers were added to boost CNN performance. The MNL network achieved 98.64% accuracy in classifying three classes case. Fast Fourier Transform (FFT) and PCA neural network (PCANet) are utilized by Li and Chen [14] as feature extraction method. From EEG signal, a frequency matrix was created using FFT, and the feature was extracted using PCANet. The extracted feature is classified using SVM. The proposed method achieves its best accuracy in classifying three classes case with a 99.6% accuracy score.

III. Material and Method

A. EEG Dataset

The data used is data that has been collected by the Department of Epileptology University of Bonn (UoB), Germany, which was obtained by Andrzejak et al. [15]. This data consists of an analog signal which is converted to 12-bit digital and then filtered by a bandpass filter in the range of 0.53 to 40 Hz. The datasets are grouped into five sets denoted by A, B, C, D, and E, where each class has different characteristics, as detailed in Table I. Each dataset has 100 single-channel EEG data segments, and each data has a duration of 23, 6 seconds for a total of 4097 samples. Sample signal for five datasets as shown in Fig. 2. This study evaluated the proposed automatic epilepsy detection using EEG waves from all sets.

B. Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) is a technique for performing signal analysis that provides a representation of a signal in time and a signal that can be computed efficiently. In DWT, the signal to be analyzed will pass through the filter process with different frequencies and scales. DWT will divide the signal into two: high frequency using a highpass filter and low frequency using a lowpass filter. DWT has a more flexible frequency window function than CWT, where the DWT frequency window narrows when observing high-frequency information and widens when analyzing low-frequency resolution. As defined in equation 1, for example, the parameter m is an integer that controls the dilation of the wavelet, the parameter m, k are an integer that controls the translation of the wavelet, s_0 is a preset scaling parameter, and its value is greater than 1. 0 is a translation parameter that has a value greater than zero and is the parent of the wavelet.

$$\Psi_{m.k}(t) = \frac{1}{\sqrt{s_0^m}} \Psi\left(\frac{t-kt_0 s_0^m}{s_0^m}\right) \tag{1}$$

In this study, DWT is used to decompose the EEG data obtained through the discrete Fourier transform process. Various wavelet families and wavelet levels are used in DWT to provide a scaling function. The wavelet decomposition L-level determines the signal's frequency band according to its level. The output of DWT were the coefficient vectors.

C. Discrete Fourier Transform

DFT is a technique to perform feature extraction using the frequency domain. DFT is beneficial because DFT makes it possible to find the spectrum of a signal with a finite duration. Since DFT treats the data periodically, it will express the input data's periodicity along with each periodic component's relative strength. The method proposed in this study uses the implementation of Fast Fourier Transform (FFT) since FFT is an efficient algorithm to compute DFT. The use of FFT is to divide the original EEG data into five frequency sub-sections, namely gamma (>30 Hz), beta (between 12 and 30Hz), alpha (between 8 and 12 Hz), theta (between 4 and 8 Hz), and delta (< 4Hz) using DFT as described in equation 2.

TABLE I. Dataset Overview

Dataset	Patient Status	Setup	Phase
A	Healthy	Surface EEG	Open Eyes
B	Healthy	Surface EEG	Close Eyes
C	Epilepsy	Intracranial EEG	Interictal Hippocampal Position
D	Epilepsy	Intracranial EEG	Interictal Epileptogenic Zone
E	Epilepsy	Intracranial EEG	Ictal

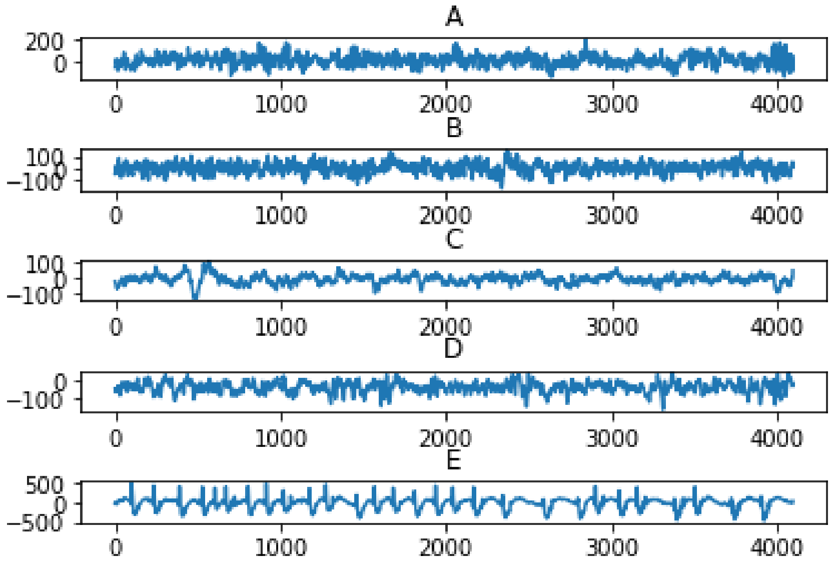


Fig. 2. Dataset EEG Samples.

$$F_n = \sum_{k=0}^{N-1} f_k \, e^{-\frac{2\pi ink}{n}}, \; n \in [0, N-1] \quad (2)$$

The transformed data is filtered using a band-pass filter to produce $F_\gamma(n), F_\beta(n), F_\alpha(n), F_\theta(n)$, and $F_\delta(n)$ signals at the respective sub-band frequencies. The EEG signal that has been in the form of frequency will be transformed again into the time domain using Inverse DFT to get an EEG signal that has been decomposed in the time domain as described in equation 3.

$$f_s(k) = \frac{1}{N}\sum_{n=0}^{N-1} F_s(n) e^{\frac{2\pi ink}{n}}, n \in [0, N-1], s = \gamma, \beta, \alpha, \theta, \delta \quad (3)$$

D. Statistical Feature

Information on data distribution can be obtained from percentiles by dividing the data into 100 equal parts. To get the percentile of p^{th}, the elements of the coefficient vector are ordered from smallest to largest. Eq (4) is used to get the n^th index of the p^{th} percentile, and N represents the length of the vector coefficient.

$$n = \frac{p}{100}(N+1) \quad (4)$$

The result was five statistical signal features retrieved from the results of each DWT coefficient vector, namely, the 95th percentile, 75th percentile, 50th percentile, 25th percentile, and 5th percentile. Thus, 5(L+1) statistical features were retrieved from all coefficient vectors of the DWT L-level decomposition result.

E. Crossing Frequency Features

Zero-Crossing Frequency (ZCF) is a condition where the two elements of the vector coefficient have a frequency that crosses zero or changes signs from positive to negative and vice versa. ZCF was chosen to replace Zero Crossing Rate (ZCR) because ZCF has a more straightforward calculation and the exact duration data. This work extracted ZCF from the coefficient vector of DWT results to capture the signal's frequency information. Suppose N is the span of the coefficient vector, $v(k)$ is the k^{th} component of the coefficient vector, and sgn is the sign function, then ZCF can be obtained by equation 5.

$$ZCF = \frac{1}{2}\sum_{k-1}^{N-1}\left|sgn\big(v(k+1)\big) - sgn\big(v(k)\big)\right| \quad (5)$$

This study also uses the Mean Crossing Frequency (MCF) to complete the signal frequency information that ZCF has obtained. MCF is described as the frequency of two subsequent components of the vector cross m; if m is the average value of the coefficient vector, MCF can be calculated by using equation 6.

$$MCF = \frac{1}{2}\sum_{k-1}^{N-1}|sgn(v(k+1) - m) - sgn(v(k) - m)| \quad (6)$$

With the L-level of decomposition, a total of 2(L+1) crossing frequency features were obtained from the coefficient vector of the DWT result.

F. Feature Selection using Genetic Algorithm

Genetic Algorithm (GA) is an optimization algorithm that is a population-based search algorithm that uses the concept of survival of the fittest. GA is inspired by natural selection[16]. A new population is generated by repeated iterations of the genetic operator on the individuals present in the population. The critical elements of GA are chromosome representation, selection, crossover, mutation, and computation of fitness functions. The Fitness function determines the ability to compete of an individual. The fitness value determines the probability of selecting an individual for reproduction. The selection phase is to choose parents based on their fitness values to carry their genes to the next generation. Crossover is the phase where the parents reproduce, and the crossover point is chosen randomly from the parents' genes. Offspring is made by exchanging genes between parents. Of the many offspring made, several offspring can experience mutations with a low random probability. This happens to maintain diversity in the population and prevent premature convergence.

GA dynamically changes the search process through crossover and mutation probabilities to achieve the optimal solution. GA has better global search capabilities because GA can modify the encoded gene. Besides that, GA can also evaluate many individuals and generate several optimal solutions. As stated by Katoch et al. [17], offspring derived from crosses of parental chromosomes have a high probability of deleting the genetic scheme of parental chromosomes. The cross formula is defined as equation (7).

$$R = \frac{G + 2\sqrt{g}}{3G} \quad (7)$$

G is the fixed number of generations determined by the population and g is the number of generations.

A classification with many features will increase the complexity of the training process. Many features also do not always result in good classification [18]. In this study, GA is used as a feature selection method to eliminate features that will not be used in the classification. Firstly, feature selection technique using GA was began by randomly creating the initial population of chromosomes, which are binary mask vectors of length comparable to the number of features. The genes on the chromosomes could take on the value 0 or 1. If the value of the i^th gene was 0, then the i^th feature was disregarded for classification; else, the feature was chosen. Feature selection using A fitness function was used to determine the quality of each chromosome. The fitness function for feature selection using GA makes use of the accuracy rate of a classifier that has been trained with chromosome-specific features. Until the final requirements are reached, the population is iteratively modified through crossover, mutation, and selection.

G. Gradient Boosting Machine Fusion

To construct new base-learners to be maximally correlated with the negative gradient of the loss function, which is related to the entire ensemble, is the principle of the Gradient Boosting Machine (GBM) [19]. Unlike the Decision Tree and Random Forest algorithms, the random forest combines several decision tree outputs to generate predictions. In GBM, each decision tree predicts from the previous error decision tree [20]. Therefore, GBM is a classification method that always tries to reduce errors. If $y = z(s(t))$ is an estimate of functional dependence, then for the loss function model, $\Psi(y, z)$, is formulated in eqution (8).

$$\hat{z}\big(s(t)\big) = \hat{y} = \arg min \Psi(y, z) \quad (8)$$

To optimize the function, $\hat{y}$ is used as a parameter in the function as $in\, \hat{y} = \sum_{i=1}^{M} \hat{y}_i$. This is what distinguishes GBM from other machine learning. In GBM, a "greedy stagewise" approach is derived from the weak-learners increment function. The function is formulated in equation (9).

$$(p_t, \theta_t) = \arg \overset{min}{p, \theta} \sum_{i=1}^{N} \Psi\big(y_i\, , \widehat{f_{t-1}}\big) + ph(x_i, \theta) \quad (9)$$

In this study, GBM Fusion is used to classify multi-class models. To improve the classification results where several classifiers would be trained as basic classifiers to classify EEG signals into five and two classes. After classifying five classes and two classes, the best results will be taken through majority voting using equation (10), suppose C is a class.

$$C(x_o) = \underset{k \in \{0,1,2,3,4\}}{\arg max} \left(\sum_{i=1}^{4} I_k \begin{matrix} s \in \{0, 1\} \\ or\ s \in \{0, 2\} \\ or\ s \in \{1, 2\} \\ or\ s \in \{1, 3\} \\ \cdots \\ or\ s \in \{3, 4\} \end{matrix} \right) \sum_{i=1}^{4} I_k\big(y_s^i\big) \quad (10)$$

H. Enhanced GBM Fusion

This study enhances the research of Sunaryono et al. [4] by evaluating the pattern of errors in the majority voting to improve the classification performance of automatic epilepsy detection in five classes case. The steps below were utilized to train Enhanced GBMs fusion and predict the class label for unknown data using Enhanced GBMs fusion.

- Obtain the decomposed signals $f_\gamma, f_\beta, f_\alpha, f_\theta$, and f_δ by decomposing the original EEG signal using DFT.
- Perform DWT with L-level decomposition to the initial EEG data to produce the coefficient vectors C1.
- Perform DWT with L-Level of decomposition to the decomposed EEG data for frequency sub-band $\alpha, \beta, \gamma, \delta$ and θ to obtain the coefficient vectors as C_2.
- Obtain feature sets F_1 by extracting 2(L+1) crossing frequency features and 5(L+1) statistical features from the coefficient vectors C_1.
- Obtain feature sets F_2 by extracting 10(L+1) crossing frequency features and 2 5(L+1) statistical features from the coefficient vectors C_2.
- Perform feature selection using GA to F_1 and F_2 to determine which are the most important features.
- Train two 5-class GBMs, X1 and X2, utilizing the selected features from F_1 and F_2 as input features, respectively.
- Train twenty 2-class GBMs with the selected features from F_1 and F_2 to classify the EEG signal as either class 0 and class 1 (named F_1^{01} and F_2^{01}), class 0 and class 2 (named F_1^{02} and F_2^{02}), class 0 and class 3 (named F_1^{03} and F_2^{03}), class 0 and class 4 (named F_1^{04} and F_2^{04}), class 1 and class 2 (named F_1^{12} and F_2^{12}), class 1 and class 3 (named F_1^{13} and F_2^{13}), class 1 and class 4 (named F_1^{14} and F_2^{14}), class 2 and class 3 (named F_1^{23} and F_2^{23}), class 2 and class 4 (named F_1^{24} and F_2^{24}), class 3 and class 4 (named F_1^{34} and F_2^{34}).
- Suppose x_0 is an EEG signal without labels. Using F_1 and F_2, predict the class label of x_0 to obtain y_1 and y_2, respectively.

a) If y_i = 0, then predict the class label of x_0 using models F_i^{01}, F_i^{02}, F_i^{03}, and F_i^{04} to obtain $y_i\, y_i^{01}, y_i^{02}, y_i^{03}$, and y_i^{04} from each model, for i = 1, 2.

b) If y_i = 1, then predict the class label of x_0 using models F_i^{01}, F_i^{12}, F_i^{13}, and F_i^{14} to obtain $y_i\, y_i^{01}, y_i^{12}, y_i^{13}$, and y_i^{14} from each model, for i = 1, 2.

c) If y_i = 2, then predict the class label of x_0 using models F_i^{02}, F_i^{12}, F_i^{23}, and F_i^{24} to obtain $y_i\, y_i^{02}, y_i^{12}, y_i^{23}$, and y_i^{24} from each model, for i = 1, 2.

d) If y_i = 3, then predict the class label of x_0 using models F_i^{03}, F_i^{13}, F_i^{23}, and F_i^{34} to obtain $y_i\, y_i^{03}, y_i^{13}, y_i^{23}$, and y_i^{34} from each model, for i = 1, 2.

e) If y_i = 4, then predict the class label of x_0 using model F_i^{04}, F_i^{14}, F_i^{24}, and F to obtain $y_i\, y_i^{04}, y_i^{14}, y_i^{24}$, and y_i^{34} from each model, for i = 1, 2.

- Class x_0 was predicted by majority vote based on the result of 5-class and 2-class GBMs.
- Evaluate the pattern of errors from GBMs Fusion majority voting to disregard the pattern of errors.
- Class x_0 predicted using majority voting on the result of 5-class and 2-class GBMs but while disregarding the error pattern.

Suppose x_0 is some EEG signals that must be classified after the feature extraction and selection stage. x_0 is classified using two 5-class GBMs, and suppose the result is class 4 from F_1 and class 4 from F_2. Further classification with twenty 2-class GBM, because both results from F_1 and F_2 is class 4 then the model that will give output only the model F_1^{04}, F_1^{14}, F_1^{24}, F_1^{34}, F_2^{04}, F_2^{14}, F_2^{24}, and F_2^{34}, the given output is class 4, class 4, class 2, class 4, class 4, class 4, class 2, and class 4, respectively. The final prediction of class x_0 after majority voting is class 4. However, because x_0 has the original class 2, the prediction result of x_0 is wrong. Therefore, the 8 GBMs models that give results of class 4 are ignored so that the majority voting results become class 2.

I. Experimental Setup

Experiments for this study have been conducted to validate the proposed method for detecting epilepsy using EEG signals in three cases. In the first case, two 5-class GBMs were trained with F_1 and F_2 to classify EEG signals. In the second case, twenty 2-class GBMs were trained with F_1 and F_2 to classify EEG into two classes, namely class 0-1, class 0-2, class 1-2, class 1-3, class 2-4, class 0-3, class 0-4, class 1-4, class 2-3, and class 3-4. In the third case, Enhanced GBM Fusion was utilized to classify EEG signals into five classes, and the results will be compared to those of previous studies.

This experiment was run on a mid-spec computer with a specification of 2.2GHz Intel(R) Core(TM) i7-8750H, 16GB RAM, NVIDIA GeForce GTX 1050 Ti GPU, and Windows 10 Home Single operating system to ensure that the proposed method is implemented on everyday life. The proposed method uses the python programming language with several libraries, namely NumPy [21], DEAP [22], scikit-learn [23], and PyWavelets [24]. This experiment uses fold-cross validation in which the EEG data is randomly divided into ten sections with equal proportions for each section.

IV. Result and Discussion

A. Classification of Five Class GBM

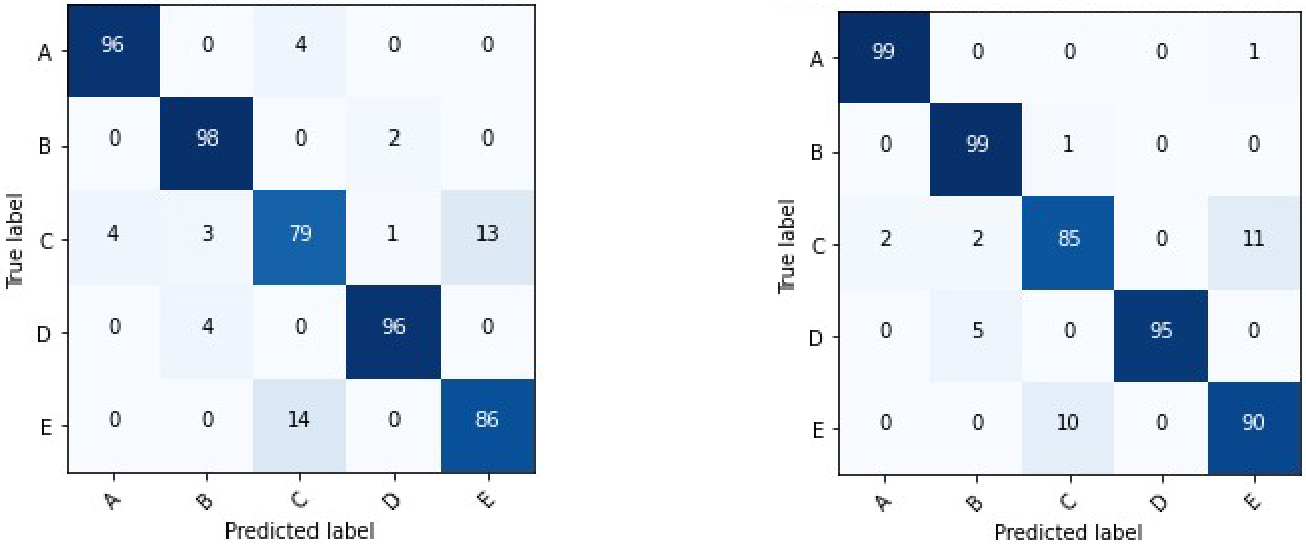
The accuracy of the 5-class classification using GBM with F_1 and F_2 has been summarized in TABLE II. The classification results using F_1 have a higher average of 91.13%, compared to F_2 with an average of 88.08%. The best result using F_1 feature is using Daubechies 6 (Db6) family wavelet with a decomposition level of 8 with the accuracy score of 91.99%. However, Db6 with a decomposition level of 8 needs more selected features to achieve the best accuracy than Symlet 16 (Sym16) and Symlet 20 (Sym20). For F_2 feature, the best result achieved is 93.6% by applying Biorthogonal 5.5 (Bior5.5) with the decomposition level of 3 while having the least original feature and having the lowest decomposition level compared to the rest as in Table II. The F_2 data has the same number of features after feature selection, while in F_1 data, more selected features are needed to get the highest accuracy results.

As shown in Fig. 3, confusion matrix was utilized to evaluate EEG signals that were incorrectly classified by F_1 using Db6 wavelet with decomposition level 8 and F_2 using bior5.5 wavelet with decomposition level 3. 5 classes model A-B-C-D-E FFT-Sub-band-Wavelet had a better classification result in class A, B, C, and E, while from 5 classes model A-B-C-D-E Wavelet had a better classification in class D.

The experimental results showed that raising the decomposition level of DWT does not necessarily increase the classification accuracy of 5-class GBM. These results also demonstrated that conducting DFT as added feature extraction method prior to DWT to generate features set F_2 does not always increase the accuracy of classification for 5-class GBM.

TABLE II. Summary of Five Class GBM Classification

Classifier	***Wavelet***		***Original Feature***	***Selected Feature***	***Accuracy with selected Feature***
	Family	***Level***			
F_1 Classifier	Db6	8	63	29	91.99
	Sym16	5	42	21	90.39
	Sym20	5	42	24	91.00
F_2 Classifier	Bior5.5	3	140	65	93.60
	Symll	5	210	65	89.20
	Db15	7	280	65	87.80



(a) Five Classes Model a-b-c-d-e.

(b) Five Classes Model -a-b-c-d-e Fft-Subband-Wavelet.

Fig. 3. Confusion Matrices of 5-Class GBM using each Features.

B. Classification of Two Class GBM

The result of the classification of 2-class GBM using F_1 and F_2 features have been summarized in Table III and Table IV, respectively. The wavelet family and decomposition level used in F_1 and F_2 are the same for each class classifier. In the experiment before using the feature selection, the accuracy results using F_2 data have higher accuracy results with an accuracy of 95.08% compared to F_1 data with an average of 93.99%, as in TABLE III and TABLE IV. The only F_1 feature before feature selection with higher accuracy against F_2 is using F_1^{14} classifier using rbio2.4 wavelet family and decomposition level of 6 with an accuracy score of 90.73% against F_2^{14} with an accuracy of 90.04% by applying the same wavelet family and decomposition level. In the experiment using feature subset, the average accuracy of the F_1 and F_2 features increased with an average accuracy of 99.19% in both data, with the highest accuracy being 100% and the lowest being 94%, as shown in TABLE III and TABLE IV, respectively. These results demonstrated that using feature selection most of the time can improve classification accuracy. The experimental results also showed that conducting frequency sub-band decomposition prior to DWT to generate features set F_2 can improve the classification accuracy of 2-class GBM.

C. Enhanced GBM Fusion Classification

Previously, the results of the 2-class model accuracy were obtained using F_1 and F_2, where the results from these models were combined and used in GBM Fusion and Enhanced GBM Fusion using majority voting. The classification results using GBMs Fusion resulted in an accuracy of 97.2% in the classification of 5 classes, with 14 misclassifications listed in Table V. Table V shows that the misclassification results mainly occur in the data with the highest majority voting value of 3, 5, or 8. This value will be used as a value of 0 during the enhanced GBM Fusion classification. The accuracy of 99.8% in the 5-class classification was also successfully achieved using Enhanced GBM Fusion by leaving one misclassification result at index 440 with confusion matrix as in Fig.4. This happened because the majority voting result in Enhanced GBM fusion at index 440 was 3 in the 5th row, which means the prediction result in the index 440 is class 4, which should be class 0. As shown in Fig. 4, all the EEG signals from class B, class C, class D, and class E were correctly predicted by Enhanced GBMs Fusion. Only one signal from class A was misclassified as a class E.

TABLE III. SUMMARY OF TWO CLASS GBM CLASSIFICATION USING F_1

Classifier	Family	Level	Original Feature	Selected Feature	Accuracy with Original Feature (%)	Accuracy with Selected Feature (%)
F_1^{01}	Db24	2	21	4	95,12	100
F_1^{02}	Db5	4	175	78	93,87	98,5
F_1^{12}	Db38	1	14	7	90,54	100
F_1^{13}	Db10	2	105	63	98,03	99,49
F_1^{24}	Sym15	5	210	106	98,46	94
F_1^{03}	Db1	1	14	6	96,98	100
F_1^{04}	Db18	2	21	12	92,5	100
F_1^{14}	Rbio2.4	6	49	15	90,73	100
F_1^{23}	Bior1.3	4	35	14	91,28	100
F_1^{34}	Coif14	4	35	11	92,46	100
Average Accuracy (%)					93.99	99.19

TABLE IV. SUMMARY OF TWO CLASS GBM CLASSIFICATION USING F_2

Classifier	Family	Level	Original Feature	Selected Feature	Accuracy with Original Feature (%)	Accuracy with Selected Feature (%)
F_2^{01}	Db24	2	21	4	95,65	100
F_2^{02}	Db5	4	175	78	94,84	98,5
F_2^{12}	Db38	1	14	7	92,96	100
F_2^{13}	Db10	2	105	63	99	99,49
F_2^{24}	Sym15	5	210	106	98,99	94
F_2^{03}	Db1	1	14	6	98,98	100
F_2^{04}	Db18	2	21	12	92,71	100
F_2^{14}	Rbio2.4	6	49	15	90,04	100
F_2^{23}	Bior1.3	4	35	14	94,68	100
F_2^{34}	Coif14	4	35	11	93	100
Average Accuracy (%)					95.08	99.19

TABLE V. CLASSIFICATION ERROR FROM GBM FUSION

Index	**Model**	**Expected Result**	**Classification Result**
37	[0 0 2 0 8]	2	4
150	[8 0 2 0 0]	2	0
165	[8 0 2 0 0]	2	0
192	[0 0 2 0 8]	2	4
213	[0 0 8 0 2]	4	2
292	[0 8 2 0 0]	2	1
370	[0 8 0 2 0]	3	1
397	[0 8 2 0 0]	2	1
426	[0 8 0 2 0]	3	1
437	[0 0 8 0 2]	4	2
440	[2 0 5 0 3]	0	2
452	[0 0 8 0 2]	4	2
458	[0 0 8 0 2]	4	2
487	[0 8 0 2 0]	3	1

Table VI is comparison proposed method with the results of previous studies, the highest accuracy of five classes classification by the previous studies is 97.39% by Sunaryono et al. [4], which means that this study has an increase in the five-class classification by 2.41% from the best results in previous studies. The proposed method has the potential to classify five classes for the detection of epilepsy. However, this method is limited with the dataset that is used. The effect of this limitation that is this method still lack validation from other datasets.

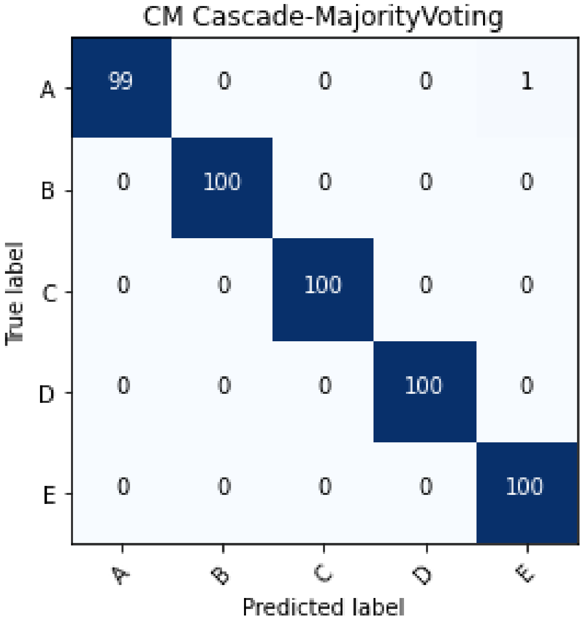


Fig. 4. Confusion Matrix of Enhanced GBM Fusion.

TABLE VI. COMPARISON TABLE WITH PREVIOUS STUDIES

Dataset	**Method**	**Study**	**Accuracy (%)**	**This Study Accuracy (%)**
A-B	GBM Fusion	[4]	**100**	**100**
	CNN+Scalogram	[7]	95.5	
A-C	GBM Fusion	[4]	**100**	98.5
	CNN+Scalogram	[7]	96.5	
A-D	GBM Fusion	[4]	**100**	**100**
	CNN+Scalogram	[7]	**100**	
A-E	MNL Network	[13]	99.52	**100**
	CNN+Scalogram	[7]	99.5	
	Symlet Wavelets, PCA, GBM-GSO	[8]	**100**	
	GBM Fusion	[4]	**100**	
	DWT, Fuzzy Approximate Entropy, SVML	[25]	**100**	
	FFT-based PCANet, SVM	[14]	**100**	
	RCMDE, SVM	[12]	**100**	
B-C	GBM Fusion	[4]	**100**	**100**
	CNN+Scalogram	[7]	99	
B-D	GBM Fusion	[4]	**100**	99.49
	CNN+Scalogram	[7]	**100**	
B-E	GBM Fusion	[4]	**100**	**100**
	Symlet Wavelets, and PCA, GBM-GSO	[8]	**100**	
	CNN+Scalogram	[7]	**100**	

	FFT-based PCANet, SVM	[14]	**100**	
	MNL Network	[13]	99.11	
C-D	GBM Fusion	[4]	94	**100**
	CNN+Scalogram	[7]	85.71	
C-E	MNL Network	[13]	98.02	98.99
	Symlet Waveletsand PCA, GBM-GSO	[8]	98.4	
	GBM Fusion	[4]	**100**	
	CNN+Scalogram	[7]	98.50	
	FFT-based PCANet, SVM	[14]	**100**	
D-E	FFT-based PCANet, SVM	[14]	99	**100**
	Symlet Wavelets, and PCA, GBM-GSO	[8]	98.1	
	LMD+GA+SVM	[26]	98.1	
	GBM Fusion	[4]	99.49	
	CNN+Scalogram	[7]	98.50	
	MNL Network	[13]	97.63	
	MDE, SVM	[12]	96.5	
A-B-C-D-E	MEMD + ANN	[27]	87.2	**99.8**
	GBM Fusion	[4]	97.39	
	ToC + DNN	[28]	97.2	
	CNN+Scalogram	[7]	93.60	
	SeizureNet	[11]	95.84	
	DWT-SVD-F*k*NN	[9]	93.33	
	1-D-DNN	[10]	93.55	
	MNL Network	[13]	93.55	

V. Conclusion

In this study, Enhanced GBM Fusion is proposed to be an automatic epilepsy detection method from EEG signal data. The proposed method obtains an accuracy value of 99.8% in classifying five classes A-B-C-D-E on a dataset from the University of Bonn. EEG signal data were decomposed using DWT and DFT as feature extraction methods. The decomposed signal is used to extract the crossing frequency feature and statistical feature. Genetic Algorithm is used as a feature selection method to get discriminatory features to improve the classification performance. For the whole experiment, the proposed method can improve the accuracy compared to normal GBM in classifying EEG signals. With the results that have been obtained, this study can be a reference for the medical world to detect epilepsy automatically so that people with epilepsy can be diagnosed accurately and quickly.

The drawback of the proposed method is that the determination of patterns on Enhanced GBM Fusion to improve performance must be done by hard coding. The performance of the proposed method may not be comparable to that of other datasets.

Acknowledgment

This study is a follow up research from Sunaryono et al. [4] with the title of “Gradient boosting machines fusion for automatic epilepsy detection from EEG signals based on wavelet features”.

This research was funded by the Ministry of Research, Technology and Higher Education of the Republic of Indonesia under Penelitian Disertasi Doktor (PDD) Program, and under scholarship scheme No. T/927/IT2/HK.00.01/2021 managed by Institut Teknologi Sepuluh Nopember (ITS) Surabaya.

References

[1] “Epilepsy.” https://www.who.int/news-room/fact-sheets/detail/epilepsy (accessed Jun. 04, 2022).

[2] Satyender, S. K. Dhull, and K. K. Singh, “A Review on Automatic Epilepsy Detection from EEG Signals,” Lecture Notes in Electrical Engineering, vol. 668, pp. 1441–1454, 2021, doi: 10.1007/978-981-15-5341-7_110.

[3] S. Supriya, S. Siuly, H. Wang, and Y. Zhang, “Automated epilepsy detection techniques from electroencephalogram signals: a review study,” Health Information Science and Systems, vol. 8, no. 1, p. 33, 2020, doi: 10.1007/s13755-020-00129-1.

[4] D. Sunaryono, R. Sarno, and J. Siswantoro, “Gradient boosting machines fusion for automatic epilepsy detection from EEG signals based on wavelet features,” Journal of King Saud University - Computer and Information Sciences, 2021, doi: 10.1016/j.jksuci.2021.11.015.

[5] J. Wu, T. Zhou, and T. Li, “Detecting Epileptic Seizures in EEG Signals with Complementary Ensemble Empirical Mode Decomposition and Extreme Gradient Boosting,” Entropy 2020, Vol. 22, Page 140, vol. 22, no. 2, p. 140, Jan. 2020, doi: 10.3390/E22020140.

[6] I. Ullah, M. Hussain, E. ul H. Qazi, and H. Aboalsamh, “An automated system for epilepsy detection using EEG brain signals based on deep learning approach,” Expert Systems with Applications, vol. 107, pp. 61–71, Oct. 2018, doi: 10.1016/J.ESWA.2018.04.021.

[7] Ö. Türk and M. S. Özerdem, “Epilepsy Detection by Using Scalogram Based Convolutional Neural Network from EEG Signals,” Brain Sciences

2019, Vol. 9, Page 115, vol. 9, no. 5, p. 115, May 2019, doi: 10.3390/BRAINSCI9050115.

[8] X. Wang, G. Gong, and N. Li, "Automated Recognition of Epileptic EEG States Using a Combination of Symlet Wavelet Processing, Gradient Boosting Machine, and Grid Search Optimizer," Sensors 2019, Vol. 19, Page 219, vol. 19, no. 2, p. 219, Jan. 2019, doi: 10.3390/S19020219.

[9] N. Singh and S. Dehuri, "Multiclass classification of EEG signal for epilepsy detection using DWT based SVD and fuzzy kNN classifier," Intelligent Decision Technologies, vol. 14, no. 2, pp. 239–252, Jan. 2020, doi: 10.3233/IDT-190043.

[10] W. Zhao et al., "A Novel Deep Neural Network for Robust Detection of Seizures Using EEG Signals," Computational and Mathematical Methods in Medicine, vol. 2020, 2020, doi: 10.1155/2020/9689821.

[11] W. Zhao and W. Wang, "SeizureNet: a model for robust detection of epileptic seizures based on convolutional neural network," Cognitive Computation and Systems, vol. 2, no. 3, pp. 119–124, Sep. 2020, doi: 10.1049/CCS.2020.0011.

[12] Sukriti, M. Chakraborty, and D. Mitra, "Automated detection of epileptic seizures using multiscale and refined composite multiscale dispersion entropy," Chaos, Solitons & Fractals, vol. 146, p. 110939, May 2021, doi: 10.1016/J.CHAOS.2021.110939.

[13] G. Zhang et al., "MNL-Network: A Multi-Scale Non-local Network for Epilepsy Detection From EEG Signals," Frontiers in Neuroscience, vol. 14, p. 870, Nov. 2020, doi: 10.3389/FNINS.2020.00870/BIBTEX.

[14] M. Li and W. Chen, "FFT-based deep feature learning method for EEG classification," Biomedical Signal Processing and Control, vol. 66, p. 102492, Apr. 2021, doi: 10.1016/J.BSPC.2021.102492.

[15] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Physical Review E, vol. 64, no. 6, p. 061907, Nov. 2001, doi: 10.1103/PhysRevE.64.061907.

[16] Z. Michalewicz, C. Z. Janikow, and J. B. Krawczyk, "A modified genetic algorithm for optimal control problems," Computers & Mathematics with Applications, vol. 23, no. 12, pp. 83–94, 1992, doi: https://doi.org/10.1016/0898-1221(92)90094-X.

[17] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," Multimedia Tools and Applications, vol. 80, no. 5, pp. 8091–8126, Feb. 2021, doi: 10.1007/S11042-020-10139-6/FIGURES/8.

[18] J. Siswantoro, H. Arwoko, and M. Z. F. N. Siswantoro, "Fruits Classification from Image using MPEG-7 Visual Descriptors and Extreme Learning Machine," 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020, pp. 682–687, Dec. 2020, doi: 10.1109/ISRITI51436.2020.9315523.

[19] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," Frontiers in Neurorobotics, vol. 7, no. DEC, p. 21, 2013, doi: 10.3389/FNBOT.2013.00021/BIBTEX.

[20] V. K. Ayyadevara, "Gradient Boosting Machine," Pro Machine Learning Algorithms, pp. 117–134, 2018, doi: 10.1007/978-1-4842-3564-5_6.

[21] C. R. Harris et al., "Array programming with NumPy," Nature 2020 585:7825, vol. 585, no. 7825, pp. 357–362, Sep. 2020, doi: 10.1038/s41586-020-2649-2.

[22] F.-A. Fortin, U. Marc-André Gardner, M. Parizeau, and C. Gagné, "DEAP: Evolutionary Algorithms Made Easy François-Michel De Rainville," Journal of Machine Learning Research, vol. 13, pp. 2171–2175, 2012, Accessed: Jun. 04, 2022. [Online]. Available: http://deap.gel.ulaval.ca,

[23] F. Pedregosa FABIANPEDREGOSA et al., "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011, Accessed: Jun. 04, 2022. [Online]. Available: http://scikit-learn.sourceforge.net.

[24] G. R. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, "PyWavelets: A Python package for wavelet analysis," Journal of Open Source Software, vol. 4, no. 36, p. 1237, Apr. 2019, doi: 10.21105/JOSS.01237.

[25] Y. Kumar, M. L. Dewal, and R. S. Anand, "Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine," Neurocomputing, vol. 133, pp. 271–279, Jun. 2014, doi: 10.1016/J.NEUCOM.2013.11.009.

[26] T. Zhang and W. Chen, "LMD Based Features for the Automatic Seizure Detection of EEG Signals Using SVM," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 8, pp. 1100–1108, Aug. 2017, doi: 10.1109/TNSRE.2016.2611601.

[27] A. Zahra, N. Kanwal, N. ur Rehman, S. Ehsan, and K. D. McDonald-Maier, "Seizure detection from EEG signals using Multivariate Empirical Mode Decomposition," Computers in Biology and Medicine, vol. 88, pp. 132–141, Sep. 2017, doi: 10.1016/J.COMPBIOMED.2017.07.010.

[28] R. Sharma, R. B. Pachori, and P. Sircar, "Seizures classification based on higher order statistics and deep neural network," Biomedical Signal Processing and Control, vol. 59, p. 101921, May 2020, doi: 10.1016/J.BSPC.2020.101921.

Intelligent Framework for Enhancing the Quality of Online Exams based on Students' Personalization

Ayman E. khedr[1], Abdulwahab Ali Almazroi[2], Amira M. Idrees[3]
University of Jeddah, College of Computing and Information Technology at Khulais[1]
Department of Information Systems, Jeddah, Saudi Arabia[1]
University of Jeddah, College of Computing and Information Technology at Khulais[2]
Department of Information Technology, Jeddah, Saudi Arabia[2]
Faculty of Computers and Information Technology, Future University in Egypt, Egypt[3]

***Abstract*—In education sector, personalization is an evolutionary term that gained a high attention due to its effectiveness in raising the enterprise competence level. This research aims at proposing a novel model for effective smart testing, which considers the student's Facebook activities in determining the students' personality and constructing his suitable exam. The aim of this examination perspectives to ensure the reliable student evaluation according to his gained knowledge to ensure that no other factor interferes which may negatively affect the reliable evaluation. The research also applies text analytics techniques to ensure the exam balance. The proposed model has been applied and evaluated with professors' percentage equal to 96.5 % and successfully reach students satisfaction percentage with average equal to 96.63%.**

Keywords—Personalization; data mining; sentiment analysis; social networks; e-learning

I. Introduction

Online education is no longer considered as an additional tool, but a vital solution that strongly supported the whole educational operation against crashing. Although this situation was not planned, however, results revealed that online education is the main factor for increasing information retention, reducing learning time, and raising the stakeholders' satisfaction level which confirmed that this adaption in the learning strategies is here to stay. The transformation of the education system towards online education has been already introduced with limits, however, this situation has moved to a significant surge forward. This shift away from classrooms does not only include the teaching classes, but it also requires other learning activities adaptation such as projects' work, practical sessions, as well as the immense need to adapt the student's assessment perspective [1]. Consequently, bolstering the online learning systems' capabilities can be currently considered a strategic objective for the educational field.

Moreover, online education has raised the alarm to focus on personalized learning as the pieces of evidence have confirmed its effectiveness of tuning the evaluation process to rely on students' personalization rather than memorization [2]. As highlighted in [3], personalized learning supports students to perceive learning with enthusiasm as it becomes more relevant and compatible with the student's characteristics [4]. Accordingly, focusing on exams personalization has become one of the goals that gained a high priority and importance as it could lead to a strong positive impact. Such an opportunistic assessment perspective in conducting the students' exams in an adequate environment based on their own pace ensures equal opportunity for all students [5]. This increasing goal for performance evaluation based on individual characteristics has been globally monitored. This recent evaluation vision has gone far beyond the traditional approach with changing the whole evaluation scheme. This research targets to intelligently transform the evaluation process from stressing, unaccommodating, and monotonic process to an effective, knowledgeable, and reliable perspective, which could be tagged as "smart testing".

Data as well as information analytics techniques play a vital role in smart testing that could successfully incorporate the personalization process into smart testing by the intervention of the behavioral students' data for tracing the suitable exam and facilitating the questions' characteristics selection [6]. On the other hand, data resources such as social networks, forums, and corporates' databases are the foundations for the personalization process for monitoring the personal behavior and characterize the student personality which consequently supports the smart testing process [7]. These resources are the pillar for revealing the student's primitive actions, his interaction preferences, as well as his trends. This revealed information feeds the testing process with the required measuring criteria for generating a suitable exam for the student [8].

This research proposes a smart model for generating a suitable exam for the student based on his personality. The student's personality has been determined based on his social network activities and educational data. The proposed model depends on the corporate database and the Facebook data as the main data resources for the personalized smart testing process.

II. Related Work

Enhancing the online education towards a more reliable process has been proposed in several researches [9] [10]. One of these researches have been presented by [11] who developed interactive learning resources with proposing to considers the learner activities during the learning process, however, the proposed enhancement focused on the closed activities cycle which was later considered by [12] could mislead the explored actions by not considering the traditional student behavior. Additionally, the research by [13] focused on tracking the student skills targeting the learning path recommendation based on a set of conducted exams. however, this approach lacked considering many exam criteria during the evaluation process such as the difficulty level.

Engaging intelligent systems in the e-learning process has been introduced from many perspectives. Different researches proposed intelligent methods for learning objectives recommendations with respect to the student progress [14]. Another research by [15] adopted the recommendation perspective to recommend the suitable department for the Fayoum university students. Other researchers have applied a variety of intelligent techniques such as the Bayes model [16], evolutionary algorithms [17], and association network [18]. In [19], Item Response Theory has been applied targeting to update the education path for the student according to his ability. Additionally, a recent research by [20] proposed a recommender system for lessons' study plan to maintain the student learning time. The proposed researches focused on the learner path; however, it is vital for an intelligent assessment method which opens the adaptation perspective using reliable criteria.

Focusing on the contribution of text analytics in the e-learning process. Text analytics has contributed to many researches in different tasks. A research in [21] proposed investigating the capability of the student in managing the exam time during the exam using text analytics, however, no consideration of other exam characteristics such as the exam level or specialty. Another research in [22] introduced text analytics in assignments' assessment for detecting plagiarism level.

Although as mentioned by [23], focusing on the student's social networking activities for discovering his behavior patterns is still considered an unexplored field, however, some researchers proposed some directions. While different empirical researches have investigated the impact of social networking and the education sector stakeholders' performance [24] [25], others have focused on the reasons that raise the power of social networking on the students in the learning process [26] [27]. It is a fact that social network influence on the educational process can take either direction, positive or negative [28], however, the success lies in lighting the shed towards the correct path [29]. Although some of the studies have tried to highlight the negative impact of the time spent in using social networking and the student's performance [30] [31], however, most of these studies revealed that there is no significant relationship between the two activities [32] [33]. Moreover, social networking extended its intervention on the learning process by offering academic resources and encourage the direct interaction among the education sector stakeholders [34]. On the other hand, researchers have highlighted that social networking users who are underage, therefore most probably students, have limited their social networking relations to their friends, not their family. Focusing on the interaction methods, some researchers argued that girls are most probably using images and share more videos on their posts more than boys [35]. More recently, the researches in [36] [37] [38] have proposed the approach of considering the Facebook activities to support the learning process by active discussion, sharing material, and continuous communication. Other researches such as in [39] [40] focused on the relation between students' engagement in Facebook activities and their willingness for class engagement. The presented researches and others have shyly considered social networking in the learning process with a limited role for continuous communication, while in the current research, social networking has a vital role and proved its effectiveness in enhancing the learning process reliability.

III. The Proposed Model for Smart Testing

The proposed smart testing model includes three main stages, "social-based exploration for the student preference", "semantic-based question bank construction", and "setting the student preference-based exam". The following sections will discuss each stage in detail.

A. *Stage One: Social-based Exploration for the Student Preference*

This stage aims at identifying the suitable type of questions for each student. Three types are available, they are illustration, long text, and short text questions. In this stage, the students' main data is collected from the faculty database including name, id, gender, email, education history (place, courses, and degrees), and Facebook account. The student's current educational status is already monitored including the registered courses, the number of credit hours gained, and remaining credit hours.

Social networks are considered the main data source in this stage. Using the Facebook account, the student's public posts are extracted and categorized to be either text or other media. This research only focuses on these two categories as the aim of this classification is to identify the suitable questions' type for the student. Moreover, the text category is also classified as short or long text. The short text category leads to the closed test questions including MCQ and T/F, while the long text leads to the open text questions including the discussion and state questions. It is vital to mention that the research depends on the student's social public data in order to avoid violating the student data privacy as social network users.

The student social activities data source is represented as follows:

The set of student's posts is represented by a set of vectors, each vector includes the post identified and its content.

$\text{ST\_Posts (STID)} = \{<\text{Post\_ID, PContent}>_j \mid j \in N\}$

All the students' posts are represented in the parent set as follows:

$\text{ALL\_ST\_Posts} = \bigcup \text{ST\_Posts (STID)} = \{<\text{STID, ST\_Posts (STID)}>\}$

Each post belongs to one of the five types, they are: short text, long text, illustration, long text &illustration, and short text & illustration. Detecting the post type is performed and the tagging is applied which can be represented as follows:

$\text{ST\_TagPosts (STID)} = \{<\text{Post\_ID, PContent, PType}>_j \mid j \in N\}$

The posts' types are then weighted to determine the student preferences, the performed steps are as follows:

The percentage of the posts with type "short text" referring to all number of posts

$$\text{PText}\ (STID)\ = \frac{\sum_{n=1}^{k} \text{PText(STID)}}{|\text{ST\_TagPosts (STID)}|} * 100$$

PText(STID) ⊆ ST_TagPosts (STID) where PType = Short Text

The percentage of the posts with type "long text" referring to all number of posts

$$\text{PText}\ (STID)\ = \frac{\sum_{n=1}^{k} \text{PText(STID)}}{|\text{ST\_TagPosts (STID)}|} * 100$$

PText(STID) ⊆ ST_TagPosts (STID) where PType = Long Text

The percentage of the posts with type "fig" referring to all number of posts

$$\text{PFig (STID)}\ = \frac{\sum_{n=1}^{k} \text{PFig(STID)}}{|\text{ST\_TagPosts (STID)}|} * 100$$

PFig(STID) ⊆ ST_TagPosts (STID) where PType = Illustration

Then the student preferences are then arranged in descending order according to the revealed percentage. The student preference set has three ordered members, each member is a vector that includes the preference order, preference type, and percentage.

Skill (STID) = {<order, preference, Percentage> | order ∈ {1,2,3}, preference ∈ {short text, long text, illustration}}. In the case of equal percentages, then the student will be considered according to the equal percentages by stating the included types in the hybrid set. Although the proposed model considers the highest student's preference, however, ordering the student's skill is vital in case that the first preference does not match with the suitable type with the course questions' type. In this case, the following student preference is considered.

B. Stage Two: Semantic-based Question Bank Construction

This stage focuses on building the test bank for the subject under examination. The main source of this stage is the course curriculum. The following steps describe the construction process in detail. In this stage, a semantic-based mesh tree is built relating the questions' tags with the subject key terms according to the subject nature. Each key subject can be related to one or more question tags. Both keys; key terms and key question tags; are tagged with one of the three types; text, illustration, and hybrid; which highlight the possible type of questions for these keys. The generated tree is used for the examination of the test bank questions' suitability which is an additional step to ensure test bank stability. Additionally, the test bank questions are tagged with the suitable keywords targeting to build an ILOs based balanced exam.

1) Step 1: Key questions tags determination: The outcome of this step is determining the question tags that are included in the exams. the questions' tags set is built based on Bloom's Taxonomy Verb Chart [41]. Bloom Taxonomy has 271 verbs which are listed as Key question terms. These terms are tagged with the suitable category and skill which highlight that this term examines a certain skill and suitable for a certain category. Bloom's Taxonomy can be represented as a set of vectors, each vector includes the verb and the corresponding skill.

Bloom's Taxonomy = {<V_j, S_j >}

Where j ∊ N

V_j is one of the Bloom verbs

S_j is the corresponding key skill, S_j ∊ Skill, Skill = {knowledge and understanding, intellectual, practical}

The key questions' tags set is represented as follows:

KQues = {<T_i, S_i, C_i>}

Where i ∊ N

T_i is a verb in the Bloom set which represents the question key

S_i is the corresponding key skill, S_i ∊ Skill

C_i is the corresponding category, C_i ∊ Category, Category = {short text, long text, figure}

2) Step 2: Key terms extraction: The main source to extract the questions' key terms is the course curriculum. The course curriculum which includes the Intended learning objects (ILOs) which describe the required topics for the course. For example, "Identify the principles of economics and management" is one of the ILOs of the database systems course. Analyzing this sentence to identify the main keywords leads to extracting a set of four keywords "identify, principles, economics, management". ILOs are analyzed to extract the main keywords of the course subject. Additionally, the course recommended textbook is also considered to expand the keywords' set. The recommended textbook is considered to match the extracted keywords with the book subjects' headings with also considering the subheadings as part of the main heading. The relation between the course ILOs and the textbook headings is the pillar of building a semantic network that fully considers the related course keywords.

Formally describing the tagged sentences is as follows:

$$\text{ILOTag}\ (S_i) = \{< K_1, Tag_1 >, < K_2, Tag_2 >, \ldots < K_n, Tag_n >\}$$

Where K_n is the token and tag_n is its attached tag

Extracting the subject key terms with the construction of the semantic network is presented by the following algorithmic steps.

// ILOs key terms semantic relations

```
For each ILO in the course description
        Extract main terms
        Apply terms' tagging
        Identify verbs as the key question tags
Identify nouns as the key terms
Build key terms semantic relations (In-relation) where In-relation identify the terms in the same ILO
// Textbook key terms semantic relations
Extract headings' tree
For each heading level
   For each heading
        Extract main terms
        Apply terms' tagging
        Identify verbs as the key question tags
Identify nouns as the key terms
Build Level 1_Heading semantic relations (In-relation) where In-relation identify the terms in the same heading
Build Level 2_Heading semantic relations (parent-relation) where parent-relation identify the relation between headings' terms
// integrate Textbook-ILO network
For each key term in the ILO, identify synonym and antonym
Extract matched heading key term with ILO key term, synonym, and antonym
Integrate the two networks' branches
// Build Key terms/Key question tags semantic relation
For each key term in the key terms set
Identify synonym and antonym
Apply KeyTerm relation set
For each key question tag in the question tags set
Identify synonym and antonym
Apply QuesTag relation set
For each ILO
Identify ILO key term as $K_i$
Extract KeyTerm $K_i$ relation set
Identify ILO verb as $QT_i$
Identify question tag $QT_i$ relation set
Append Key terms/Key question tags semantic relation ( $K_i$ – $QT_i$ )
```

3) Step 3: Building semantic based test bank: The test bank is a group of question sets. each set includes the questions that examine one of the course ILOs. In this step, building the test bank question sets is achieved. Each question in the test bank is tagged with an attribute vector. The vector includes time, mark, difficulty level, associated key terms set, associated key question tags set. The time, mark, difficulty level attributes are identified by the professor, while the associated key terms and questions tags are explored by applying the following algorithmic steps.

```
For each question in the question bank
        Identify question key tags
        Identify question key terms
        Match question key terms with network members
        If success
Tag the question with network members id
                Identify the question ILO id
                add question to ILO questions' set
        Else
                Raise an alarm
```

At the end of this step, each ILO id questions' set should have a group of members which are the questions that examine the ILO. If the ILO set is empty, then acquiring questions from the professor is applied.

C. Stage Three: Setting the Student Preference-based Exam

At this stage, the student exam is built according to his revealed skill. The questions are selected randomly from the corresponding categories based on the required number of questions and the determined exam time. The following algorithm steps are performed to build the required exam.

```
Given   Student_Pref (STID)
        Test_Bank = {<Pref, Skill, Difficulty, {< QID, QText, QDuration, QMark >}}
Set Exam_time
Set Exam_Grade
If Student_Pref (STID) ∊ {ShortText, LongText, illustration}
Questions_Set = select questions from Test_Bank where Category = Student_Pref (STID)
Else
Questions_Set = select questions from Test_Bank
Initiate Exam_test bank
Set courseSkillsSet
SkillTime = ExamTime / |courseSkillsSet|
SkillGrade = ExamGrade / |courseSkillsSet|
Set CourseDiffSet
While ALLQues_time < Exam_Time  && ALLQues_Grade < Exam_Grade {
For each skill j in course_skills
    Set ILO set
    ILOTime = Skilltime / | ILO set |
    ILOGrade = SkillGrade / |ILO set|
    DiffTime = ILOTime / | CourseDiffSet|
    DiffGrade = ILOGrade / | CourseDiffSet|
    For Each ilo in ILO set
      For each Difficulty i in course_Difficulty
While DiffQues_time < DiffTime &&  DiffQues_grade < DiffGrade {
Question = Random Choice (Questions_test bank) where Skill = j and Difficulty = I and   ILO = ilo
 If DiffQues_Time + Question.Question_Duration <= Diff_Time
&& DiffQues_Grade+ Question.Question_Grade <= DiffGrade
 {
   DiffQues_Time = DiffQues_Time + Question.Question_Duration
   DiffQues_Grade = DiffQues_Grade + Question.Question_Grade
   Append (Exam_Content, Question_Text)
   AllQuesTime = AllQuesTime + DiffQues_Time
   AllQuesGrade = AllQuesGrade + DiffQues_Grade }}}
```

IV. Experimental Case Study

The research has two validation milestones. First, validating the model success in exploring the suitable student's skill based on his social activities are performed while validating the constructed exam is also considered as the second milestone. The following sections discuss the case study setup, the performed steps, and the validation results.

A. Stage One: Exploration for the Student Preference

In this research, the dataset included the skills classification for 473 students. The dataset is based on the data which was provided in the research of [1]. The research of [1] included research that adapts the online system to satisfy the student. The original dataset included 752 students while the response was 631 students, and a satisfaction level with an average of 75 % of them have been satisfied with the proposed enhanced system with a total of 473 students in the online exam based on a series of conducted exams. The current research acquired the students' data who had been satisfied with a total of 473 students. The acquired data included all the personal and academic students' data. Extracting the required social data is performed through implementing a simple program and the data is stored as four attributes, student id, corresponding social user id, post content, post type. As a result of this process, the final experiment dataset included 380 students after removing the incomplete students' records.

A five-month range of public Facebook posts; starting from January 2018 to May 2018; have been gathered from the students' accounts. This short time range has been considered to avoid the large number of posts that may require extensive tools and software for big data which was not the scope of this paper. The posts' contents have been processed to detect the post type. The post type classification is illustrated in Table I.

The results have been compared with the results in [1]. This comparison has revealed 93% success for correct exploration to the student's most suitable skill. By analyzing the presented results, a focus has been performed on the incorrect classified students and the time period for their posts has been enlarged. A sample of these students' posts with a range of eight months starting from November 2017 to May 2018 has been extracted and the model has been applied for the student re-exploration. 40 % of these students have been re-classified to the correct skill. Therefore, this minor experiment revealed that the 7% failure in the correct classification was according to the short time range for the extracted social data due to the minor activity of the student in the three months on focus.

B. Stage Two: Semantic-based Question Bank Construction

This stage focuses on constructing the courses' questions' bank in a semantic-based format. This research argues that this proposed construction supports the applicability to generate a suitable balanced preference-based exam. The following subsections discuss the conducted steps in detail with demonstrating the required validation.

1) Step 1: Key questions tags determination: As previously discussed, the Bloom verbs are considered the question tags. 271 verbs are included in the key questions' tags set. For example, the key question term "Define" is suitable for the understanding skill and the text category. A sample of the taxonomy is illustrated in Table IV.

2) Step 2: Key terms extraction: The experiment focused on "systems analysis and design" course which are obligatory to the students in the information systems department. Table V illustrates the ILOs statistics for the courses on focus with examples for each course.

TABLE I. Post Types

Post Content	**Type**
Single Video or image	Illustration
Single text sentence where words' count is less than 10 words	Short text
Single text sentence where words' count is more than 10 words	Long text

According to this classification, the students' posts have been tagged with the appropriate class, a statistical illustration of the students' posts is presented in Table II, while another perspective of this classification was illustrated in Table III presenting the statistics of classified students based on their social activity.

TABLE II. Statistics of Posts

Type	**Statistics**
multimedia	445,584
Short text	420,254
Long text	239,901
Illustration & Short text	558,069
Illustration & Long text	168,704
Total	1,832,512

TABLE III. Statistics of Students

Type	**Number of Students (%)**
multimedia	70 (12.96 %)
Short text	62 (11.48 %)
Long text	40 (7.4 %)
Illustration & Short text	238 (44.07 %)
Illustration & Long text	130 (24.07%)
Total	540

TABLE IV. SAMPLE OF WORDNET VERBS CATEGORIES

Key Question Term	**Skill**	**Category**
write	Knowledge and Understanding	Long Text & Short Text
state	Knowledge and Understanding	Long Text & Short Text
outline	Knowledge and Understanding	Long Text & Short Text
classify	Knowledge and Understanding & intellectual	Long Text & Short Text
define	Knowledge and Understanding	Long Text
discuss	Knowledge and Understanding	Long Text
describe	Knowledge and Understanding	Long Text
identify	Knowledge and Understanding	Long Text
explain	Knowledge and Understanding	Long Text
summarize	Practical	Long Text
assess	Practical	Long Text
criticize	Practical	Long Text
differentiate	Practical	Hybrid
construct	intellectual	Hybrid
design	intellectual	Hybrid
compare	intellectual	Hybrid
solve	intellectual	Hybrid
develop	intellectual	Hybrid
apply	intellectual	Hybrid
draw	intellectual	Illustration
Sketch	Intellectual & Practical	Illustration

TABLE V. SAMPLE OF COURSES' ILOS

Course	**Skill**	**ID**	**ILOs**
Systems Analysis and Design	Knowledge and Understanding	1	Discuss specifications and strategic planning for a given project.
		2	Recognize different methods for data analysis and design.
		3	Illustrate management process for software projects and productions.
	Intellectual Skills	4	Analyze information systems problems, setting goals and requirements.
		5	Identify main ideas, patterns, components, attributes and detect relationships between these components in software analysis with different designs.
		6	Select appropriate methodologies and techniques for a given problem solution and setting out their limitations and errors
	Practical	7	Describe different analysis and design methodologies.
		8	Analyze system process and data requirements
		9	Apply different IS methodologies for analysis and design.

Text processing tasks have been applied to the courses' ILOs [42]. According to [43] the most suitable processing tasks for correctly extracting keywords that ensure avoiding conflict or missing the word's meaning are lemmatization and part of speech tagging.

Consequently, lemmatization and Part of Speech tagging (POS) are applied for the ILOs, nouns are extracted as keywords and indexed. As a result, each ILO is tagged with two references, the skill Id, and the set of corresponding keywords' Ids. For Example, ILO text "Discuss specifications and strategic planning for a given project" has the following two steps:

Lemmatization: "Discuss specification and strategic planning for a give project"

POS tagging: $\text{ILOTag}\ (ILO_i) = \{<\text{Discuss}, \text{V}><\text{specification}, \text{NNS}><\text{and}, \text{CC}>,<\text{strategic}, \text{JJ}><\text{planning}, \text{NN}><\text{for}, \text{IN}>,<\text{a}, \text{DT}><\text{give}, \text{VBN}><\text{project}, \text{NN}>\}$

The ILO is then tagged with the set of keywords' Ids which are: "specification", "planning", "project" in addition to the skill id. Moreover, synonyms and antonyms of the specified terms are identified and included in the set members. Statistically speaking, the experiment included a total of 9 ILOs, a total of the extracted keywords equal to 26. Each ILO was tagged with a range of 2 to 7 keywords. At the end of this stage, the relation between the ILOs and the corresponding keywords is revealed for each course.

3) Step 3: Building semantic based test bank: In this stage, the relation between the test bank questions and the ILOs are explored based on the extracted keywords. It is recommended that the test bank for each course included 300 questions. Each question is tagged with one of the three types (short text, long text, multimedia) according to its key question tag. Tagging each question with its corresponding keywords is then applied.

According to [44], bi-gram keywords provide the highest accurate accuracy. Therefore, uni-gram keywords and bi-gram keywords are constructed, then, the matching process is applied. Each question is tagged with two keywords subsets, the first subset includes the matched uni-gram keywords while the second includes the matched bi-gram keywords. the following example is a sample of the performed process while Table VI presents a statistical presentation with the step results.

Question: A system request will generally have these items: project sponsor; business need; business requirements; business value; special issues or constraints

Keyword (s): system, request, project, sponsor, business, need, requirement, value, issue, constraint

Matched bi-gram Keyword (s): (system, problem), (system, requirement), (requirement, problem), (problem, limitation), (problem, error), (limitation, error).

Matched ILOs ID: 4, 6, 8

Validating the tagging process has been performed by the course professors. The validation process has confirmed the higher accuracy in tagging the questions with bi-gram keywords, therefore, following the bi-gram keywords' tagging was the main path to the next phase for constructing the student's exam.

Table VII demonstrates the evaluation criteria and the evaluation percentage according to the professors' perspective. It is shown that the proposed model has succeeded in classifying the test bank questions with an average success percentage equal to 96.6 %.

TABLE VI. A STATISTICAL PRESENTATION WITH THE STEP RESULTS

	Systems analysis and design
Short text questions	150
Long text questions	94
Illustration questions	56
Average questions / ILO	33

TABLE VII. TEST BANK VALIDATION

No of questions that are classified to the correct skill(s)	582
No of questions that are classified to the suitable ILO(S)	570
No of questions that are tagged by the suitable keyword(s)	588

C. Stage Three: Setting the Student Preference-based Exam

As mentioned in [45], the stakeholders' satisfaction is one of the main objectives for organizations in any field. Therefore, this research targets not only moving in the right track for the smart testing paradigm but the students' satisfaction, as well as the professors' satisfaction, are also essential objectives for this research. Therefore, generating the preference-based exams for the students has been applied and validated by both the professors and the students as follows:

A focus on the successfully classified students is performed, fifty successfully classified students have been randomly selected for testing the stage of generating a suitable exam. The selection process ensured the variety of the students' preferences. As previously illustrated, five categories are considered, therefore, twenty of the students/preference/course have been targeted. Generating the exams followed the proposed algorithmic steps and the constructed tests have been reviewed by the course professors. Table VIII demonstrates the statistical measures of the conducted exams. It is shown that the proposed model has succeeded in generating a suitable exam for the students with an average success percentage equal to 96.5 %. The 3.5 % failure goes to the lack of questions that could be required to complete the exam time or the exam grade, therefore, the experiment has been revised after feeding the test bank with an additional 50 questions for each course that has a variety in the marks distribution and required time. This adaptation has filled the required gab and the exams were successfully generated based on the professors' evaluation.

On the other hand, the students' satisfaction has been measured and the results are demonstrated in this section. Conducting two generated exams have been developed, first a randomly generated exam without considering their preferences and a randomly generated exam after applying the proposed model. The results are presented in Table IX while illustrated in Fig. 1- 3.

TABLE VIII. EXAM VALIDATION

Total No. of generated exams	50
No. of exams that are correctly generated	48
Average No of questions / Student	25
Minimum No of questions / Student	19
Maximum No of questions / Student	40

TABLE IX. COMPARISON FOR THE AVERAGE STUDENTS' RESULTS

	Excellent		**Very Good**		**Good**		**Fair**		**Fail**	
	No	%	No	%	No	%	No	%	No	%
Random Exam	6	12	8	16	16	32	9	18	11	22
Model-Based Exam	16	32	10	20	14	28	5	10	5	10

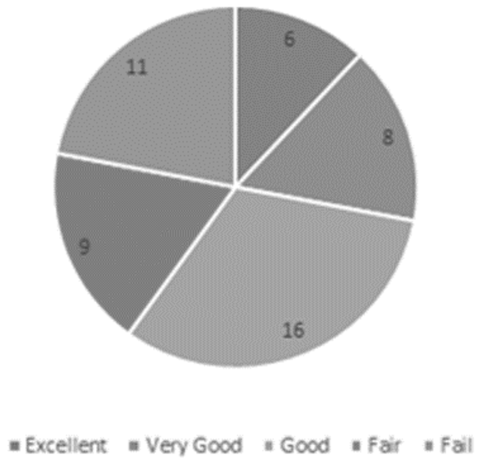


Fig. 1. Random Exam Results' Distribution.

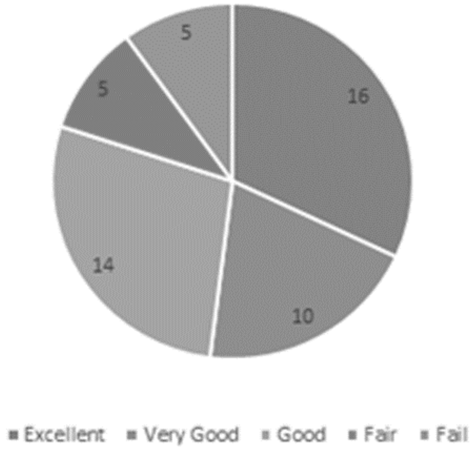


Fig. 2. Model-based Exam Distribution.

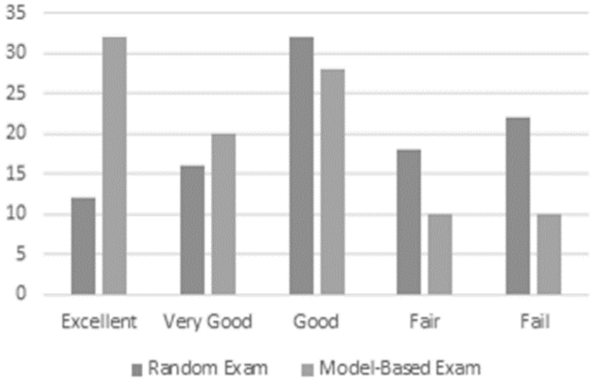


Fig. 3. Comparison between the Two Exams' Results (Random Exam and Model-based Exam).

V. Conclusion

This study proposed a student personalized exam model which is based on two main approaches. The first approach focuses on exploring the student's personalization characteristics targeting to detect his questions' preferences. This target has been on focus to ensure the questions' types suitability for the student which leads to a higher evaluation accuracy to the student learning level. The second approach utilizes text analytics techniques in different goals. It contributes to revealing the student preferences in addition to its contribution in constructing the exam contents. The proposed model considers the balance while generating the student exam according to different criteria including the difficulty level, the intended learning outcomes coverage, the required time, and the degrees' distribution in addition to the student's preferences. According to the experiment evaluation results, the proposed model succeeds in reaching 96.5% of the professor satisfaction in the constructed exams.

The proposed model confirmed its effectiveness in approaching smart testing, however, enhancement directions can further contribute effectively to the same direction. One of the proposed enhancements is considering the students' emotions which provide more accuracy to his on-time preferences. Another direction is the automated generation to the questions' banks to ensure the full course coverage. Finally, the current research relied on the highest student preference, considering the preferences according to its adequacy level could be an effective enhancement to ensure the full smart testing perspective coverage.

Acknowledgement

This work was funded by the University of Jeddah, Saudi Arabia, under Grant No. (UJ-20-092-DR). The authors, therefore, acknowledge the technical and financial support by the University.

References

[1] A. E. Khedr and A. M. Idrees, "Adapting Load Balancing Techniques for Improving the Performance of e-Learning Educational Process," Journal of Computers, vol. 12, no. 3, pp. 250-257, 2017.

[2] D. H. A. Hassouna, A. E. Khedr, A. M. Idrees and A. I. ElSeddawy, "Intelligent Personalized System for Enhancing the Quality of Learning," Journal of Theoretical and Applied Information Technology, vol. 98, no. 13, pp. 2199-2213, 2020.

[3] D. Nandigam, S. S. Tirumala and N. Baghaei, "Personalized learning: Current status and potential," in 2014 IEEE Conference on e-Learning, e-Management and e-Services (IC3e), 2014.

[4] A. E. Khedr, . S. A. Kholeif and S. H. Hessen, "Adoption of cloud computing framework in higher education to enhance educational process," International Journal of Innovative Research in Computer Science and Technology (IJIRCST), Volume 3, Issue 3, pp. 150 - 156, 2015.

[5] A. E. Khedr and A. I. El Seddawy, "A Proposed Data Mining Framework for Higher Education System," International Journal of Computer Applications, vol. 113, no. 7, pp. 24-31, 2015.

[6] E. Afify, A. Sharaf Eldin, A. E. Khedr and F. K. Alsheref, "User-Generated Content (UGC) Credibility on Social Media Using Sentiment Classification," FCI-H Informatics Bulletin, vol. 1, no. 1, pp. 1-19, 2019.

[7] A. E. Khedr, A. M. Idrees and F. K. Alsheref, "A Proposed Framework to Explore Semantic Relations for Learning Process Management," International Journal of e-Collaboration, vol. 15, no. 4, 2019.

[8] Hegazy, Abdel Fatah; Khedr, Ayman E.; Al Geddawy, Yasser;, "An Adaptive Framework for Applying Cloud Computing In Virtual Learning Environment at Education aCase Study of"AASTMT"," in International Conference on Communication, Management and Information Technology (ICCMIT 2015), 2015.

[9] S. H. Lee, "Learning vocabulary through e-book reading of young children with various," Reading and Writing, vol. 30, no. 7, p. 1595–1616, 2017.

[10] J. Li, F. Ma, Y. Wang, R. Lan, Y. Zhang and X. Dai, "Pre-school children’s behavioral patterns and performances in learning numerical operations with a situation-based interactive e-book," Interactive Learning Environments, pp. 1-18, 2019.

[11] G. J. Hwang and C. L. Lai, "Facilitating and Bridging Out-Of-Class and In-Class Learning: An Interactive E-Book-Based Flipped Learning Approach for Math Courses," Journal of Educational Technology & Society, vol. 20, no. 1, 2017.

[12] K. Mouri, N. Uosaki and H. Ogata, "Learning analytics for supporting seamless language learning using e-book with ubiquitous learning system," Journal of Educational Technology & Society, vol. 21, no. 2, pp. 150-163, 2018.

[13] A. Mostafa, A. E. Khedr and A. Abdo, "Advising Approach to Enhance Students' Performance Level in Higher Education Environments," Journal of Computer Science, vol. 13, no. 5, pp. 130-139, 2017.

[14] K. Mouri, Z. Ren, N. Uosaki and C. Yin, "Analyzing Learning Patterns Based on Log Data from Digital Textbooks," International Journal of Distance Education Technologies (IJDET), vol. 17, no. 1, pp. 1-14, 2019.

[15] A. M. Idrees and M. H. Ibrahim, "A Proposed Framework Targeting the Enhancement of Students’ Performance in Fayoum University," International Journal of Scientific & Engineering Research, vol. 9, no. 11, 2018.

[16] A. H. Nabizadeh, A. Mário Jorge and J. Paulo Leal, "Rutico: Recommending successful learning paths under time constraints," Adjunct publication of the 25th conference on user modeling, adaptation and personalization, p. 153–158, 2017.

[17] K. Govindarajan and V. S. Kumar, "Dynamic learning path prediction-a learning analytics solution," Technology for education (T4E), 2016 IEEE eighth, pp. 188-193, 2016.

[18] F. Yang, F. Li and R. Lau, "Learning path construction based on association link network," Advances in web-based learning-ICWL, p. 120–131, 2012.

[19] X. An and Y. F. Yung, "Item response theory: what it is and how you can use the IRT procedure to apply it," SAS Institute Inc. SAS364-2014, vol. 10, no. 4, 2014.

[20] A. H. Nabizadeh, D. Gonçalves, S. Gama, J. Jorge and H. N. Rafsanjani, "Adaptive learning path recommender approach using auxiliary learning objects," Computers & Education, vol. 147, 2020.

[21] Y. Levy and M. M. Ramim, "A Study of Online Exams Procrastination Using Data Analytics Techniques," Interdisciplinary Journal of E-Learning and Learning Objects, vol. 8, 2012.

[22] G. Akçapınar, "How automated feedback through text mining changes plagiaristic behavior in online assignments," Computers & Education, vol. 87, 2015.

[23] H. Zarzour, S. Bendjaballah and H. Harirche, "Exploring the behavioral patterns of students learning with a Facebook-based e-book approach," Computers & Education, vol. Accepted June 2020, 2020.

[24] E. Alwagait, B. Shahzad and S. Alim, "Impact of social media usage on students academic performance in Saudi Arabia," Computers in Human Behavior, vol. 51, pp. 1092-1097.

[25] H. Hawi and M. Samaha, "To excel or not to excel: Strong evidence on the adverse effect of smartphone addiction on academic performance," Computers and Education, vol. 98, p. 81–89, 2016.

[26] Q. Li, "Characteristics and social impact of the use of social media by Chinese Drama," Telematics Inform, vol. 34, no. 3, p. 797–810, 2017.

[27] M. Samaha and N. Hawi, "Associations between screen media parenting practices and children’s screen time in Lebanon," Telematics Inform, vol. 34, p. 351–358, 2017.

[28] S. Madden, M. Janoske and R. L. Briones, "The double-edged crisis: Invisible Children’s social media response to the Kony 2012 campaign," Public Relations Review, vol. 42, no. 1, pp. 38-48, 2016.

[29] S. Lovea, M. Sanders, K. Turner, M. Maurange, T. Knott, R. Prinzc, C. Metzler and A. Ainsworth, "Social media and gamification: Engaging vulnerable parents in an online evidence-based parenting program," Child Abuse & Neglect, vol. 53, pp. 95-107, 2016.

[30] R. Gafni and M. Deri, "Costs and Benefits of Facebook for Undergraduate Students," Interdisciplinary Journal of Information, Knowledge, and Management, vol. 7, 2012.

[31] A. J. Ndaku, "Impact of Social Media on the Students’ Academic Performance: A Study of Negussie, N. and Ketema, G. (2014). Relationship between," sian J. Hum. Soc. Sci., vol. 2, no. 2, pp. 1-7, 2014.

[32] M. Koutamanis, H. Vossen and P. M. Valkenburg, "Adolescents’ comments in social media: Why do adolescents receive negative feedback and who is most at risk?," Computers in Human Behavior, vol. 53, pp. 486-494, 2015.

[33] B. Miller, A. Stewart, J. Schrimsher, D. Peeples and P. Buckley, "How connected are people with schizophrenia? Cell phone, computer, email, and social media use," Psychiatry Res., vol. 225, p. 458–463, 2015.

[34] L. Tomczyk and K. Kopecky, "Children and youth safety on the internet: experiences from Czech Republic and Poland," Telematics Inform., vol. 33, p. 822–833, 2016.

[35] E. Tartari, "Benefits and risks of children and adolescents using social media," European Scientific Journal, vol. 11, no. 13, p. 321–332, 2015.

[36] W. Peeters, "The peer interaction process on Facebook: a social network analysis of learners’ online conversations," Education and Information Technologies, vol. 24, no. 5, pp. 3177-3204, 2019.

[37] S. Toker and M. H. Baturay, "What foresees college students’ tendency to use facebook for diverse educational purposes?," International Journal of Educational Technology in Higher Education, vol. 16, no. 1, p. 9, 2019.

[38] S. Xue and D. Churchill, "A review of empirical studies of affordances and development of a framework for educational adoption of mobile social media," Educational Technology Research and Development, vol. 67, no. 5, pp. 1231-1257, 2019.

[39] Y. Hong and L. Gardner, "Undergraduates’ perception and engagement in Facebook learning groups," British Journal of Educational Technology, vol. 50, no. 4, pp. 1831-1845, 2019.

[40] N. Sheeran and D. J. Cummings, "An examination of the relationship between Facebook groups attached to university courses and student engagement," Higher Education, vol. 76, no. 6, 2018.

[41] M. T. Chandio, S. Pandhiani and R. Iqbal, "Article Bloom's Taxonomy: Improving Assessment and Teaching-Learning Process," Journal of Education and Educational Development, vol. 3, no. 2, 2017.

[42] A. E. Khedr and A. M. Idrees, "Enhanced e-Learning System for e-Courses Based on Cloud Computing," Journal of Computers, vol. 12, no. 1, 2017.

[43] M. Othman, H. Hassan, R. Moawad and A. M. Idrees, "Using NLP Approach for Opinion Types Classifier," Journal of Computers, vol. 11, no. 5, 2016.

[44] M. Othman, H. Hassan, R. Moawad and A. M. Idrees, "A linguistic approach for opinionated documents summary," Future Computing and Informatics Journal, vol. 3, no. 2, pp. 152-158, 2018.

[45] A. Khedr, S. Kholeif and S. Hossam, "Enhanced Cloud Computing Framework to Improve the Educational Process in Higher Education: A case study of Helwan University in Egypt," International Journal Of Computers & Technology, vol. 14, no. 6, 2015.

AUTHORS’ PROFILE

Ayman E. Khedr, Professor. I am currently a professor in the university of Jeddah, College of Computing and Information Technology at Khulais. I have been the vice dean of post-graduation and research and the head of Information Systems Department in the Faculty of Computers and Information Technology, Future University in Egypt. I am a professor in the Faculty of Computers and Information, Helwan University in Egypt. I have previously worked as the general manager of Helwan E-Learning Center. My research is focused around the themes (scientific) data and model management, Data Science, Big Data, IoT, E-learning, Data Mining, Bioinformatics and Cloud Computing.

Abdulwahab Ali Almazroi, Associate Professor, received his M.Sc. and Ph.D. in Computer Science from the University of Science, Malaysia, and Flinders University, Australia, respectively. He is currently serving as an Associate Professor in the Department of Information Technology, College of Computing and Information Technology at Khulais, University of Jeddah, Saudi Arabia. His research interests include parallel computing, cloud computing, wireless communication, and data mining.

Amira M. Idrees, Professor. I'm a professor in Data Science, the head of information systems department, faculty of computers and information technology, Future University in Egypt. I have been the head of scientific departments and the vice dean of the community services and environmental development, Faculty of Computers and Information, Fayoum University. I'm currently an associate professor in the Faculty of Computers and Information Technology in Future University and the head of University Requirements Unit. My research interests include Knowledge Discovery, Text Mining, Opinion Mining, Cloud Computing, E-Learning, Software Engineering, Data Science, and Data warehousing.

Churn Prediction Analysis by Combining Machine Learning Algorithms and Best Features Exploration

Yasyn ELYUSUFI, M'hamed AIT KBIR
LIST Laboratory
STI Doctoral Studies Center
Abdelmalek Essaadi University, Morocco

***Abstract*—The market competition and the high cost of acquiring new customers have led financial organizations to focus more and more on effective customer retention strategies. Although the banking and financial sectors have low churn rates compared to other sectors, the impact on profitability related to losing a customer is comparatively high. Thereby, customer turnover management and analysis play an essential part for financial organizations in order to improve their long-term profitability. Recently, it appears that using machine learning to predict churning improves customer retention strategies. In this work, we discuss some specific machine learning models proposed in the literature that deal with this problem and compare them with some emerging models, based on Ensemble learning algorithms. As a result, we build a predictive churn approaches that look at the customer history data, check to see who is active after a certain time and then create models that identify stages where a customer can leave the concerned company service. Ensemble learning algorithms are also used to find relevant features in order to reduce their number which is of great importance when performing the training step with some classical models such us Multi-Layer Perception Neural networks. The proposed approaches can achieve up to 89% in accuracy when other research works, dealing with the same dataset, can achieve less than 86%.**

Keywords—Customer churn; prediction; machine learning

I. Introduction

Customer churn analysis deals with customer attrition rate in companies that offer some services. According to Forrester research statistics related to churn impact [1], it costs five times more to catch new customers compared to keeping the existing ones. As well, the Harvard Business School report states that on average, 5% increase in customer retention results in increase from 25% to 95% in profits. There are so many classic ways to keep customers from leaving [2]; by finding answers to the following issues:

- What are we, as a company, doing to cause customer turnover?
- What are our customers doing that is contributing to their leaving?
- How can we better manage our customer relationships to make sure it does not happen?

On the other hand, predictive analytics uses prediction models that approximate the risk of customer attrition and the degree of their dissatisfaction with regard to services offered by the organization. In addition, new technologies have increased bank access to customer data, which has made customer attrition analysis increasingly easy and accessible. In fact, predicting customer attrition can help banks to plan suitable marketing campaigns to convince clients who are potentially candidates for leaving [3].

Overall, actually demand for customer attrition analysis is increasing, the study of the characteristics related to the customers profile and behavior, by consulting their transactions history, remains the most widely used approach in research works related to this domain, most of these are statistical learning methods. This takes up the following question, which learning model can best predict customer churn? By taking a look at methods used in the literature, we can find that popular methods to predict churn likelihood are logistic regression (LR) [4], K-Nearest Neighbor (KNN) [5], decision trees (DT) [4] and SVM [6].

In this work, we seek to use and evaluate performances of new methods that can reach best results, compared to those cited above, at predicting customer churn which can give suitable responses to the following questions:

- How successful are the cited Machine learning methods in predicting customer churn?, answering this question will allow us to know which methods give the most reliable predictions based on many metrics;
- How are the relevant features according to some machine learning algorithms ?, answering this question allows us to pick up the most relevant features so as to use the training samples with reduced size vectors, which allows time saving in both training and generalization steps. Relevant features can be combined later with high accuracy machine learning algorithms to enhance performances.

However, methods cited above have reached their limits and have practical difficulties resulting in the emergence of a new generation of algorithms called Ensemble learning algorithms: Random Forest (RF), eXtreme Gradient Boosting (XGBoot) and Light Gradient Boosted Machine (LightGBM) models [7]. Light GBM model has many of XGBoost's assets, such as sparse optimization, parallel processing, regularization and bagging [8]. These algorithms belong to two families of ensemble decision tree models, Bagging and Boosting. We are also interested in this study by artificial neural networks

(ANN) based models, especially Multi-Layers neural network [22].

In this study, we are also interested by improving performances of machine learning models, related to predicting customer churn, by exploring ensemble learning algorithms capabilities to find relevant features. Indeed, the use of relevant features with some classical machine learning models is important when carrying out the training step.

This paper is a part of a series of research carried out by a team of our laboratory, interested in users profiling and related predictive analyzes [9] [10] [11]. The work is presented as follows; in the first section we present literature review of churn prediction works using machine learning models. In the second section we present data pre-processing, modeling, and comparison with some cited works. In the third section, we present our approach of best feature exploration in order to improve the quality of prediction using Multi-Layer Perceptron model. The results interpretation and conclusion comes in the last section.

II. Related Work

Today the industry is in competition with a limited number of potential customers because of the increasing saturation of the market [12]. Customers seek value-added relationships with their suppliers in order to stay loyal [13]. Companies are therefore looking for strategies to engage customers in their process in order to have a tangible idea of the required benefit. In order to improve this process, these companies are seeking to recognize customer behaviors that indicate a decline in his relationship with the company; this is established as customer relationship management (CRM). It's used by companies to efficiently assign their resources in order to maintain and improve customer relations [14], it's also used to lock onto customers churn rate. Churning specify the decrease in the consumption of a service provided by the company or the termination of that service, it can also be defined as an un-subscription. Several models have been proposed for customer behavior prediction in order to choose the auspicious time when the human resources department must give them more attention, after which incentive strategies should be followed in order to preserve this relationship and make it as profitable as possible. Many methods have been presented in the literature to determine the churn rate, they differ depending on the contractual and non-contractual nature of the work, the corresponding data structure availability and the amount of accessible transaction history. The following paragraph will be devoted to the presentation of research works related to our problematic and the promising machine learning models they use.

The results of some churn prediction models can vary on a wide spectrum; in this specific context a study of measuring the predictive accuracy of customer churn models was presented in [15]. In the same way, Breiman made the foundation for churn analysis based on classification and regression trees [16]. Upon this work, the cited author has also built additional methods by using bootstrap aggregators or bagging methods [8]. The bagging used proceeds by bootstrapping replicates of the training set, followed by creating the aggregation of predictors. In this use case Random forests introduced in [17] builds upon the previous model by introducing a new layer obtained by randomizing the bagging. In fact, a random subset of descriptors is used to expand trees, each one use a sample of the training set. It must be mentioned that RF method is known by its sensibility to data with unbalanced number of samples, which is typical to customer churn data sets, as the percentage of customers who churn is small or relatively unknown. In the same context, Xie proposes a new learning approach, called improved balanced random forests to produce better prediction results. In the present work we will assess the impact of using data re-sampling and scaling (robust-scaling method) before applying machine learning models. The following sections will be devoted to the preprocessing, modeling, evaluation, and comparison of machine learning model performances. These methods are used to help financial organizations to make decisions about the suitable strategy to prevent the customer from leaving the financial organization.

III. Our Work

A. Goal of the Work

The goal of this work is to explore whether bank's customers are about to leave the organization or not. In order to make an efficient model to predict customer churn we will use machine learning models. The event that defines the customer abandonment is the closing of the customer's bank account Fig.1

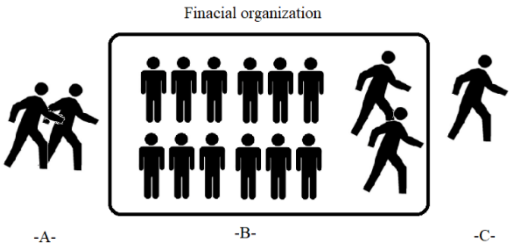


Fig. 1. Customer Churn Issue.

A-segment customers can need marketing campaigns to let them join the financial organization, when the B-segment customers can need to use machine learning models in order to predict who is about to leave and finally we can analyze C-segment customers so that more importance can be given to the suspicious profile thereby gaining more knowledge about how to make them more loyal.

B. Dataset

The dataset under study is called "Bank Customer" [18], the data was collected from an anonymous organization with customers in France, Spain and Germany. This public dataset consists of 10000 observations and 12 features, containing customer's information, where the “Exited” feature refers to customer abandonment status (target class). In order to find the churning client and to help making decision about the precaution that the financial organization should undertake we decided to study several machine-learning models, evaluate their performances and combine their decisions. The following table shows the used features, see Table I.

TABLE I. USED FEATURES

Feature	Signification
Surname	The surname of the customer
CreditScore	The credit score of the customer
Geography	The country of the customer(Germany/France/Spain)
Age	The age of the customer
Tenure	The customer's number of years in the bank
Balance	The customer's account balance
NumOfProducts	The number of bank products that the customer uses
HasCrCard	Does the customer has a card? (0=No,1=Yes)
IsActiveMember	Does the customer has an active mebership (0=No,1=Yes)
EstimatedSalary	The estimated salary of the customer
Exited	Churned or not? (0=No,1=Yes)

IV. DATA PRE-PROCESSING

Before the pre-processing, re-sampling and data scaling, we can analyze data relevance according to each feature and each machine learning algorithm. In fact, we decided to reveal the relevant feature by analyzing the columns one by one, looking for their dependency towards the churn. The following features are removed, because they do not have impact in the final decision:

- Row Number: It corresponds to the record row number
- Customer Id: It contains random values
- Surname: It represents the surname of a customer

Concerning the other features, we discuss their impact on customer churn by briefly analyzing the data.

- Credit Score: People with a credit score between 680 and 689, about 342 customers, are more loyal than others and less likely to leave the organization.
- Age: As shown in Fig. 2; it's certainly clear that the age feature is relevant, since customers with age between 35 and 55 are more likely to leave their organization than customers with other ages.

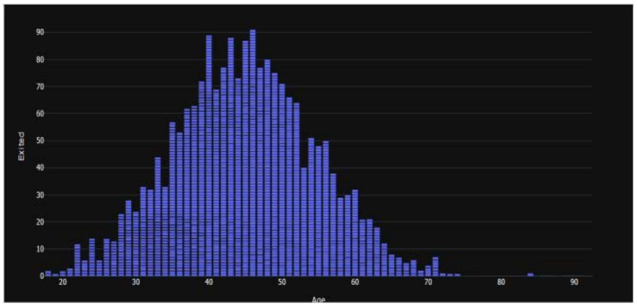


Fig. 2. Age Distribution.

- Tenure: It mentions the number of years spent by the customer with the organization. Normally, people with Tenure=7 are more loyal, about 851 customers.
- Balance: It's an important indicator, as people with a balance between 122k and 127k are more loyal to the organization.
- Number of Products: it indicates the number of products that a customer has purchased through the organization, people with Number of Products >=2 are more loyal to the organization, about 4242 customer.
- Has Credit Card: It indicates the fact that a customer has a credit card or not. People possessing a credit card are less predisposed to leave the organization, 30.09% of customers that churned have not a card, about 613 customer.
- Is Active Member: It indicates the presence of transactions over the customer account, active customers are more predisposed to stay with bank and 63.9% of customers that churned are not active.
- Estimated Salary: It indicates the estimated salary of the customer. The costumers with salary between 175k and 185k are more likely to leave the organization, about 122 costumers, while the more loyal costumer with salary between 77.5k and 82.5k, about 222 costumers.
- Gender: As shown in Fig. 3, we can infer that gender feature plays an essential role in churn prediction.
- Geography: It indicates customer's location. Being close to the bank can affect customer decision, especially with home changing.

In order to prepare data, we will apply the one hot encoding process for Geography feature to allow a more expressive categorical data, as shown in Table II.

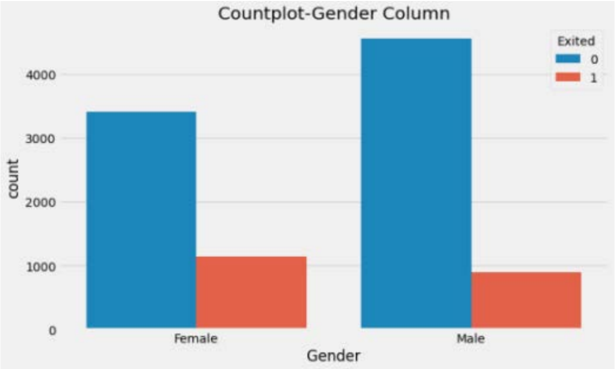


Fig. 3. Gender Distribution.

TABLE II. GEOGRAPHY FEATURE ONE HOTE ENCODING

Geography_France	Geography_Germany	Geography_Spain
1	0	0
0	0	1
1	0	0
1	0	0
0	0	1

According to Geography feature analysis, the histogram shown in Fig. 4 can infer that the total number of customers who exited is highest from Germany, which means that the bank needs to focus more on those customers followed by France customers and finally Spain customers.

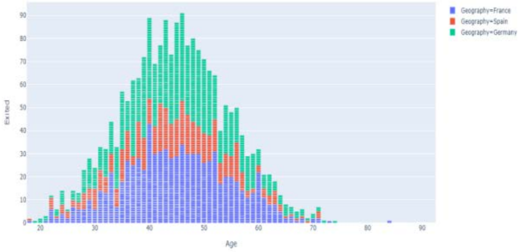


Fig. 4. Age Histogram Depending on Customer Location.

The Age feature is a numerical continuous number. According to Fig. 4, the age range between 35 and 55 are likely to leave. Furthermore, according to correlation study Fig. 5, the Age feature is correlated with 0.29 with the target and certainly will be a relevant variable for prediction.

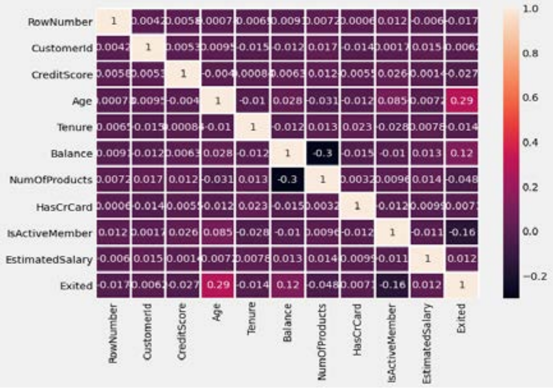
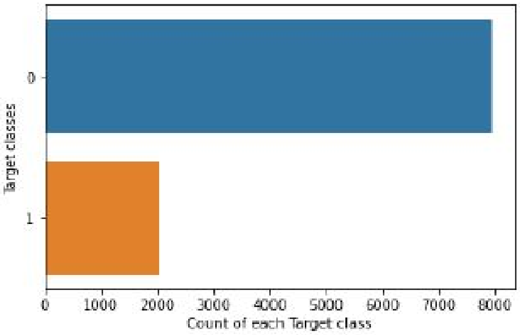


Fig. 5. Heatmap (Correlation with Target).

V. Re-Sampling And Modelling

A. Re-Sampling

In many works, the re-sampling technique is used to deal with unbalanced datasets [19] [20]. It is based on removing under-sampling respectively adding over-sampling samples from the majority in respective of the minority class. As shown in Fig. 6, we can check easily that the distribution of the target feature is unbalanced.



.

Fig. 6. Count Plots of Target Feature "Exited".

We decide to apply the over-sampling for increasing the number of samples as illustrated in Fig.7.

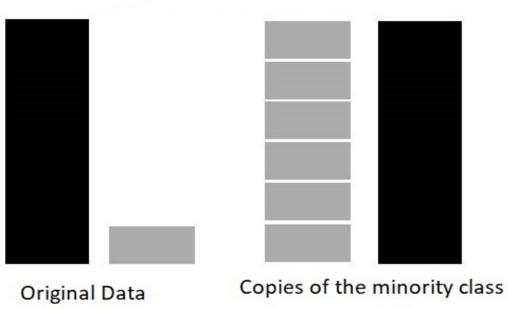


Fig. 7. Re-Sampling Method (Over-Sampling).

B. Evaluation Metrics

In order to assess machine-learning capabilities, the evaluation is based on many metrics. The most used ones are: accuracy, sensitivity, specificity, Matthew's Correlation Coefficient (MCC), that are based on confusion matrix information, presented by the following Table III, and Cohen's kappa and Matthew's correlation coefficient based on agreement and disagreement probabilities.

TABLE III. Confusion Matrix

		Actual Values	
		Success(1)	***Failure (0)***
Predicted Values	Success(1)	Truepositive	False negative
	Failure(0)	False positive	Truenegative

1) Accuracy: The accuracy calculates how many correct results your model managed to identify.

$$\text{Accuracy} = \frac{True\ positive+True\ negative}{Total\ number\ of\ samples} \quad (1)$$

2) Sensitivity: The sensitivity represents the fraction of relevant results that were retrieved; it uses also the information provided by the confusion matrix.

$$\text{Sensitivity} = \frac{number\ of\ true\ positives}{number\ of\ true\ positives+number\ of\ false\ negatives} (2)$$

3) Specifity: The specificity also called true negative rate, measures how well a model can identify true negatives.

$$\text{Specificity} = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives} (3)$$

4) Cohen's kappa: The Cohen's Kappa is a statistic measure, commonly referred to as inter-rater reliability. This metric reflects the reliability of two raters, who are rating the target result and the real one, and identifies how frequently the raters agree. Cohen's kappa, symbolized by letter $\boldsymbol{\kappa}$, this later range between -1 to 1, and it depends on the probability of agreement minus the probability of disagreement. When $\boldsymbol{\kappa}$=0 the amount of agreement has a random value, and κ=1 represents perfect agreement.

$$\%\ of\ agreement = 100 \times \kappa^2 \quad (4)$$

5) Matthew's correlation coefficient (MCC): Matthew's correlation coefficient (MCC), is a metric that aims to evaluate the quality of binary classification. MCC is often used when dataset are unbalanced, and it's unimpressed by disproportions related to dependent variables. MCC metric takes values between -1 and 1. When MCC score is equal to -1, this reflects an unexceptionable misclassification, however when it's equal to 1, a perfect classification is detected, while zero value indicates that the model is no better than random values.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}} \quad (5)$$

C. Modeling

After the application of the process of over sampling using SMOTE library of python, we count 15926 samples while the original dataset count only 10000 observations. After, we need to standardize the data using a scaling stage. We used Robust-Scaling method [21], computed by subtracting the median from the feature value and dividing by the inter-quartile range (75% value - 25% value). This step is followed by dataset splitting, into testing and training sub-sets, the training subset count 90% of dataset. In our work, we will assess the impact of seven supervised machine-learning algorithms: Logistic Regression (LR) [4], K-nearest neighbors(KNN) [5], Random Forest (RF) [22], Decision Tree [4], Support Vector Machine (SVM) [6], XGB and Light GBM [7].

In order to choose the best model, for XGB and LightGBM, the following empirical hyper-parameter setting has been adopted see Table IV:

TABLE IV. HYPER PARAMETER USED FOR XGB AND LIGHTGBM MODELS

Learning Rate	0.05
Max Dept	5
Estimators	500

The Multi-Layer Perceptron (MLP) was configured so that the input layer has a number of neurons identical to the number of features. We proceed with two hidden layers; 15 neurons are used in the first hidden layer, 10 neurons in the second hidden layer. The stochastic gradient descent (SGD) was used as the solver function, with learning rate fixed to 0.05. Finally, the model will be trained for 500 epochs.

Parameters discussed in the precedent two paragraphs are used in all our experiments.

The evaluation process of each model, before and after over-sampling process, gives metric scores indicated respectively in Table IV and Table V.

After re-sampling and scaling, the accuracy score has been enhanced for all the models except for LR, this can be justified by the fact that this technique is susceptible to over-fitting and assumes linear relationship between independent attributes.

TABLE V. EVALUATION METRIC WITH ROBUST SCALING NORMALIZATION

Name	Accuracy	Sensitivity	Specificity	Cohen's Kappa	MCC
LR	0.803	0.213270	0.960710	0.227233	0.267877
KNN	0.830	0.402844	0.944233	0.404683	0.422475
CART	0.793	0.492891	0.873257	0.370652	0.370736
RF	0.849	0.436019	0.959442	0.465812	0.489577
SVM	0.847	0.345972	0.980989	0.415727	0.470908
XGB	0.851	0.478673	0.950570	0.489586	0.504743
LightGBM	0.857	0.507109	0.950570	0.515891	0.528854
MLP	0.847	0.454976	0.951838	0.469582	0.487264

TABLE VI. EVALUATION METRIC WITH ROBUST SCALING NORMALIZATION AND (OVER-SAMPLING)

Name	Accuracy	Sensitivity	Specificity	Cohen's Kappa	MCC
LR	0.824231	0.794155	0.853598	0.227233	0.649173
KNN	0.844319	0.841169	0.847395	0.404683	0.688586
CART	0.827997	0.858958	0.797767	0.370652	0.657615
RF	0.890772	0.885642	0.895782	0.465812	0.781516
SVM	0.861896	0.822109	0.900744	0.415727	0.725570
XGB	0.888889	0.872935	0.904467	0.489586	0.777994
LightGBM	0.889517	0.874206	0.904467	0.515891	0.779227
MLP	0.863779	0.834816	0.892060	0.469582	0.728408

These results show that ensemble decision tree models and Multi-Layer neural network produce higher accuracy when dealing with the "Bank Customer" dataset. However, finding best parameters for these methods is not a small task. In fact, it is time consuming with a high number of features. Consequently, using just relevant features to present dataset samples can be helpful when setting the appropriate model parameters and can affect the classification results.

The results in table VII were obtained by research works operating on the dataset described in this paper.

TABLE VII. RESULTS OF EVALUATION METRIC USING LR, RF, AND KNN MODELS PUBLISHED IN [5]

Name	Accuracy	Sensitivity	Specificity	Cohen's Kappa	MCC
LR	0.811	0.964	0.211	0.231	0.273
RF	0.866	0.969	0.465	0.513	0.539
KNN	0.836	0.962	0.342	0.376	0.409

Results presented in this table compared to those obtained in table VI show that the accuracy was improved for LR, RF and KNN models respectively by 1.3%, 2.4% and 1.4%. On the other hand, the accuracy given by the XGB and Light GMB models was respectively increased by 2.2%, 2.3% with respect to the highest accuracy result obtained in able VII. In the following section, we will proceed to feature importance analysis, using Scikit-Learn library [23], according to RF, XGB and Light GBM, models with the best obtained accuracy score. This analysis aims to enhance the quality of prediction and to make the training phase easier for some machine learning models such as Multi-Layer neural network.

VI. FEATURE RELEVANCE EXPLORATION

Feature relevance represents the reduction in node impurity weighted by the probability of reaching that node. The number of observations cumulated by the node divided by the total number of observations corresponds to the node probability. Higher values correspond to features that are more relevant. For each decision tree, nodes relevance is calculated, by taking only two child nodes (binary tree):

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} c_{right(j)} \qquad (6)$$

ni_j: Importance of the node j

w_j: Proportion of observations that get at the node j

C_j: Impurity of node j

left (j) and right(j): Child node resulting from splitting the node j.

The relevance of each feature is then calculated by:

$$fi_i = \frac{\sum_{j: node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k \epsilon\ all\ nodes} ni_k} \qquad (7)$$

fi_i: The relevance of feature i

This term can be normalized by dividing by the sum of all feature relevance values:

$$\mathrm{normfi_j} = \frac{\mathrm{fi_j}}{\sum \mathrm{j \in all\ features\ fi_j}} \qquad (8)$$

The final feature relevance, at the Random Forest grade, is defined by the mean of all the trees. The sum of the feature's relevance value on each tree is calculated and divided by the total number of trees:

$$\mathrm{RFfi_i} = \frac{\sum \mathrm{j \in all\ trees\ norm\ fi_{ij}}}{\mathrm{T}} \qquad (9)$$

$\mathrm{RFfi_i}$: The relevance of feature is calculated from all trees in the RF model.

Norm ($\mathrm{RFfi_i}$): The normalized feature relevance for feature in the tree j.

T: Total number of trees.

The application of feature relevance calculation according to the three models with the best accuracy score > 88.8 % (RF, XGB and Light GBM) are shown in Fig. 8, Fig. 9, and Fig. 10.

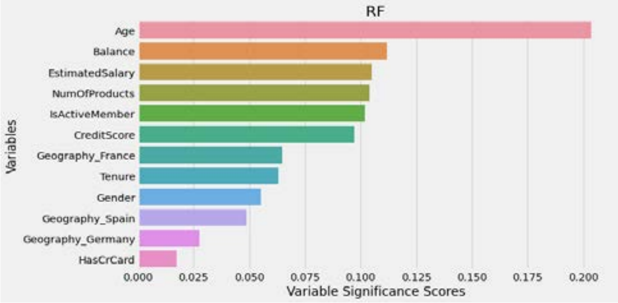


Fig. 8. RF Feature Relevance.

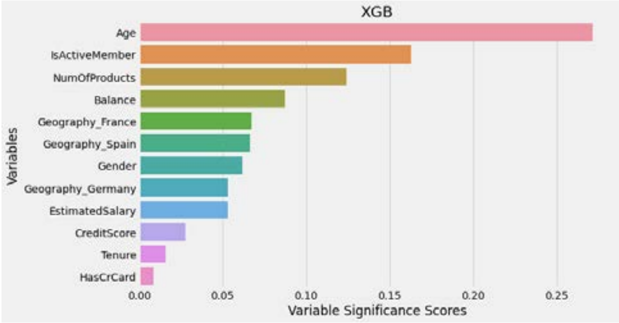


Fig. 9. XGB Feature Relevance.

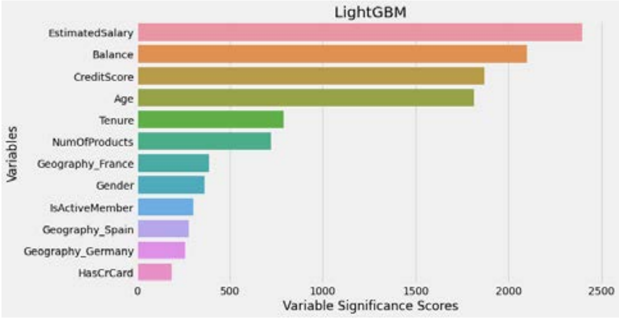


Fig. 10. Light-GBM Feature Relevance.

The feature selection process was performed using the empirical hyper-parameters setting cited below.

A. Feature Selection usingXGB and LightGBM

After the calculation of feature relevance for each machine learning algorithm, the above results show that the most eight relevant features according to RF are:

- Age,
- Balance,
- Estimated Salary,
- Number Of Products,

- Is Active Member,
- Credit Score,
- Geography_France,
- Tenure.

The most eight relevant features according to XGB are:

- Age,
- Is Active Member,
- Number Of Products,
- Balance,
- Geography_France,
- Geography_Spain,
- Gender,
- Geography_Germany.

While the most eight relevant features according to Light-GBM model are:

- Estimated Salary,
- Balance,
- Credit Score,
- Age,
- Tenure,
- Number Of Products,
- Geography_France,
- Gender.

The goal of this part of study is to find relevant features according to each machine-learning algorithm and to try after to use just those features with other algorithms such as MLP neural network. This can make the training process easier and leads to time saving in both training and generalization steps.

According to the chosen hyper-parameters, the accuracy reaches 88.88% for the XGB model and 88.95% for Light GBM model.

B. *Using MLP Model with Relevant Features*

After implementing feature scaling for all inputs variables in order to have values with comparable ranges, we will study the effect of using relevant features with MLP model. The most relevant features are those that appear at least twice in the union of relevant features sets elaborated by RF, XGB and Light-GBM models. These features are:

- Age,
- Balance,
- Estimated Salary,
- Number Of Products,
- Is Active Member,
- Credit Score,
- Geography_France,
- Tenure,
- Gender.

We did the simulation with nine relevant features with MLP model.

After applying the feature selection with ensemble decision tree models and using these features to train MLP model, only the Cohen's Kappa metric has been improved, which reflect the perfect agreement between the observed and predicted outcomes, it has been improved by 19.40%. According to the results shown in Table VIII, we can deduce that even after removing three features, the MLP performances are not drastically affected.

TABLE VIII. OBTAINED METRICS 9 MOST RELEVANT FEATURES WITH MLP MODEL

Name	**Accuracy**	**Sensitivity**	**Specificity**	**Cohen's Kappa**	**MCC**
MLP	0.831764	0.844981	0.818859	0.663601	0.663902

VII. RESULTS INTERPRETATIONS

In this research, we have compared the results of applying eight machine-learning algorithms, LR, KNN, CART, RF, SVM, XGB, Light GBM and MLP with other studies using "Churn Customer" dataset [5]. Our study aims to predict accurately customers predisposed to stay with bank and to honor their commitments. For a given model, a high accuracy would indicate that the model is able to predict the decision that a customer can make (exit the bank/ stay with the bank). After the evaluation of the eight models, we notice that RF model got the best accuracy score with 89.07%. It must be mentioned that we have adopted three pre-processing steps, which consists of data re-sampling, scaling and hyper parameter setting. Light GBM model was the second more accurate model, with an accuracy of 88.95%. The third more accurate model was the XGB model with an accuracy score of 88.88%. While Artificial Neural Network (MLP) classifier accuracy score got 86.37%.

Concerning feature number reduction, the use of relevant features as input of the MLP neural network shows that the predicting capabilities of this model were not seriously affected. In fact, the accuracy is still around 83.17%, even if three features were not used. This result is important because the training phase is easier to engage when using a reduced number of features.

VIII. CONCLUSION AND FUTURE WORK

In this paper, we proposed an approach for churn customer prediction, using a dataset with 10000 observations. After the comparison of our results to those obtained by other existing approaches [5], with respect to the used dataset, ensemble decision tree and Multi-layer neural network models have demonstrated good performances. In this work, we were also interested in improving the obtained results by exploring ensemble learning algorithms capabilities to find relevant features that can be fed to some classic machine learning models, as an example MLP neural network. The MLP model performances were not seriously affected by the reduction of features number, from 12 to 9, even some metrics were improved.

We can notice that this approach can be a motivation to make better decisions when dealing with personalized data and can also be generalized to deal with sophisticated concepts such as NLP (Natural language processing), in order to semantically analyze customer posts. We are also interested, as a perspective, in assisting MLP implementation by using Big Data tools to enhance its performances. In fact, it is possible to involve Spark libraries to enable running various machine-learning algorithms on distributed systems. Integrating these tools to this work will give the speed and the capacity to perform better results with a large amount of data.

REFERENCES

[1] S. Brinks, "Improving Customer Experience And Revenue Starts With The App Portfolio", A Forrester Consulting Thought Leadership Paper, March 2020.

[2] A. Gallo, "The value of Keeping the Right Customers", Harvard Business Review, October 29,2014.

[3] J. Ganesh,M.J Arnold & K.E Reynolds, “Understanding the Customer Base of Service Providers: An Examination of the Differences between Switchers and Stayers”. July 2000, Journal of Marketing 64(3), pp. 65-87.

[4] A. De Caigny,K. Coussement, & K.W De Bock, 2018. "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees". European Journal of Operational Research, 269(2), pp. 760-772.

[5] I.Tandan, E.Goteman, "Bank Customer Churn Prediction, a comparison between classification and evaluation method", UPPSALA University, June 4, 2020.

[6] W. Verbeke et al. 2012. "New insights into churn prediction in the Telecommunication Sector: A profit driven data mining approach". European Journal of Operational Research, 218(1), pp. 211-229.

[7] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.Y. Liu, "Light-GBM: A Highly Efficient Gradient Boosting Decision Tree", NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, December 2017. Pages 3149–3157.

[8] L. Breiman. "Bagging predictors", Kluwer Academic Publishers, Boston, Machine Learning, 24(2):123–140, Aug 1996.

[9] Y. Elyusufi, Z.Elyusufi, M. Ait Kbir "Social Networks Fake Profiles Detection Based on Acount Setting and Activity", SCA '19: Proceedings of the 4th International Conference on Smart City Applications, October 2019 Article No.: 37. pp. 1-5.

[10] Y. Elyusufi, Z., Elyusufi, M. Ait Kbir "Customer profiling using CEP architecture in a Big Data context", SCA '18: Proceedings of the 3rd International Conference on Smart City ApplicationsOctober 2018 Article No.: 64Pages 1–6.

[11] Y. Elyusufi, H.Seghiouer, M.A.Alimam, "Building profiles based on ontology for recommendation custom interfaces", 2014 International Conference on Multimedia Computing and Systems (ICMCS), 14-16 April 2014.

[12] W. J. Ferrier, G. K. Smith, and C. M. Grimm. "The role industry of competitive action in market share erosion and dethronement: A study of industry leaders and challengers". The Academy of Management Journal, 42(4):372–388, 1999.

[13] W. J. Reinartz and V. Kumar. "On the profitability of long-life customers in a non contractual setting: An empirical investigation and implications for marketing". Journal of Marketing, 64(4):17–35, 2000.

[14] K. Coussement, D. F. Benoit, and D. V. Poel. "Improved marketing decision making in a customer churn prediction context using generalized additive models". Expert Systems with Applications, 37(3):2132 – 2143,2010.

[15] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. H. Mason. "Defection detection: Measuring and understanding the predictive accuracy of customer churn models". Journal of Marketing Research, 43(2):204–211,April 2006.

[16] L. Breiman. "Classification and Regression Trees". Routledge Taylor & Francis Group, Published January 1, 1984 by Chapman and Hall/CRC, ISBN 9780412048418.

[17] L. Breiman." Random forests". Machine Learning, 45(1):5–32, Oct 2001.

[18] A. Manus “Predicting Churn for Bank Customers”, (Kaggle : October 2018): https://www.kaggle.com/adammaus/predicting-churn-for-bank-customers.

[19] O. Arbelaitz, I. Gurrutxaga, J.Muguerza, and J. María Pérez: "Applying Resampling Methods for Imbalanced Datasets to Not So Imbalanced Datasets", Springer-Verlag Berlin Heidelberg 2013, : CAEPIA 2013, LNAI 8109, pp. 111–120, 2013.

[20] F. Alahmari, "A Comparison of Re-sampling Techniques for Medical Data Using Machine Learning", Journal of Information & Knowledge Management, Vol(19), No. 1(2020).

[21] J.Hale," Scale, Standardize, or Normalize with Scikit-Learn", Towards data Science Corporation, Mars 4, 2019.

[22] Y. Xie, X. Li, E.W.T. Ngai, and W. Ying. "Customer churn prediction using improved balanced random forests". Expert Systems with Applications,36(3, Part 1):5445 – 5449, 2009.

[23] S.Ronaghan, "The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-learn and Spark", Towards data Science Corporation , May 11, 2018.

Rider Driven African Vulture Optimization with Multi Kernel Structured Text Convolutional Neural Network for Classifying e-Commerce Reviews

H. Mohamed Zakir[1]*
Department of Computer Science
Noorul Islam Centre for Higher Education
Kumaracoil, Thuckalay, India

Dr.S.Vinila Jinny[2]
School of Computer Science & Engineering
Vellore Institute of Technology
Vellore, India

Abstract—**Opinion mining is a natural language processing based on sentiment classification technique to determine the sentiment of the reviews. The major existing text Convolutional Neural Network (CNN) algorithms are derived based on 3×3 size kernels which extract ineffective review text-features and lead to less classification accuracy. Moreover, most of the traditional CNN versions output three classes such as positive, negative, and neutral as their classification results. Hence, a novel algorithm namely 'RAVO driven *Multi-Size Kernel structured Text CNN for classifying ecommerce reviews (MSK-TCNN-RAVO)*' is proposed in this work. This proposed approach utilizes five multi-size kernels ($3 \times 7, 5 \times 7, 1 \times 3, 1 \times 5, 1 \times 7$), multi-dimensional kernels (1D & 2D), and integrates varying size kernels to extract text-features effectively. In addition, the performance of multi-kernel CNN is highly enhanced by RAVO algorithm based on rider optimization. Moreover, the proposed approach is highly effective to process '*review-stop-words removal*' that decrease the complexity and time consumption of the opinion mining process. Most existing systems use single pooling operations which reduce feature map processing performance, hence, dual pooling operations (both Max and Average pooling) are employed in this research. Furthermore, it is configured to generate five classification outputs such as bad, fair, neutral, good, and excellent to support better decision-making with 95.5% accuracy. This method is evaluated using different quality metrics and five review-databases to measure the performance, and the results reveal that the proposed method outperforms the other existing review classification algorithms.**

Keywords—Natural language processing; opinion mining; convolutional neural network; text sentiment classification; ecommerce review

I. INTRODUCTION

Social media plays a very important role in almost everybody's day to day life. It allows the people to convey what they think and feel about the products in the e-commerce website. This is called an opinion or review. Online shopping is a form of e-commerce which allows consumers to directly buy goods or services from a seller over the internet [1]. Most of the e-commerce websites incorporate provisions which enable the consumer to post their opinions about the product, companies, or their experiences etc.

Online shopping is being performed by millions of users every day, as a result of this, a huge amount of reviews are being generated constantly. Handling these reviews manually to extract knowledge is a tedious task and hence, companies use classification tools to categorize the customer reviews to understand the customer's mindset. Opinion Mining is also known as Sentiment Analysis (SA) is the automated process of identifying opinions in text, and labeling them as positive, negative or neutral, based on the emotions expressed by the customer. It includes text analysis, computational linguistics to identify and extract subjective information in source materials [2].

The rest of the article is structured as follows: A summary of traditional methodologies for review classification is provided in Section II. Section III covers the proposed approach in detail with diagrammatic representations. Results and discussions are included in Section IV. The work has been concluded in Section V.

II. RELATED WORK

The existing review classification algorithms discussed in various works of literature are summarized in this section. Anam et al. [3] proposed a voting classifier (LR-SGD) model for textual-tweets classification as happy or unhappy for emotion recognition. The dataset contains a lot of contrary tweets which are used for evaluation. The weakness of this work is that, the combination of different models should be employed to increase the performance. Ren and Wu [4] proposed a sentence-based analysis model to identify investor herd behaviour. The data is taken from blue-chip stocks in Chinese stock market. The work's limitation is that it uses only a small amount of data for experimentation, arising doubt on the algorithm's efficacy.

Sasikala and Sheela [5] expressed a sentiment analysis technique called Deep learning modified neural network to process the online product reviews. The food review dataset has been taken as input for the proposed technique. The drawback is that, the keyword processing only detects the sentiment expressed in a single word, and frequently fails to provide all information needed to interpret the context. Zeeshan et al. [6] proposed an opinion mining methodology using lexicon and neural networks for classifying online movie reviews. The reviews are collected from the IMDB database. The disadvantage is that, the resultant vectors will be in larger dimension and contain a large number of null

*Corresponding Author.

values, resulting in sparse vectors. Jia et al. [7] proposed a hierarchical gated deep memory network with position-aware for aspect-based sentiment analysis. The algorithm is evaluated on laptop and restaurant datasets. The limitation of the work is that the accuracy is very low while predicting the neutral polarity.

Cheng et al. [8] proposed a multi-channel model that combines the CNN and the bidirectional gated recurrent unit network with attention mechanism for text sentiment orientation analysis. Two public datasets such as IMDB and Yelp 2015 review dataset is used to evaluate the algorithm. It does not include any syntactic structure features, hence, this model cannot be used for sentiment classification task. Dong et al. [9] proposed a CNN based on multiple convolutions and pooling for text sentiment classification. The reviews from English and Chinese datasets are used as input for evaluation. The four convlotuion operations proposed in this work perform low on English emotions.

Gupta et al. [10] proposed a feature-based supervised model to identify the extremist reviewers who target whole brand. The datasets are created by crawling reviews from Amazon website. The weakness of the work is that it requires larger training time for even smaller datasets. Madbouly et al. [11] proposed a hybrid classification approach of tweets based on user ranking for online social networks. The dataset consists of tweets with their feature used as input. Tweets were filtered manually to exclude non-English tweets which will increase the execution time of the algorithm. Bhalla and Bagga [12] proposed an RB-Bayes method based on Naive Baye's theorem for prediction to remove problem of zero likelihood. The algorithm is evaluated on a small dataset which contains text data. The efficiency of the model is not proved in large-scale databases. Zhang and Zhong et al. [13] proposed an e-commerce reviews mining method called sentiment similarity analysis to explore user's similarity and trust. The experimental dataset is collected from Amazon.com. The essential parameters involved in calculating the trust between users is not considered in this work which leads to less accuracy.

Aziz et al. [14] proposed the Contextual analysis mechanism to find the relationship between words and sources to predict Supervised machine learning model performance. The experiment is conducted on four different domain datasets collected from Amazon. The result of the prediction algorithm is less while performing real time analysis for individual changes of dataset. Iqbal et al. [15] proposed a hybrid framework for sentiment analysis which bridges the gap between lexicon-based and machine learning approaches. The reviews dataset is collected from UCI ML repository. The proposed framework works only for specific domain and does not support other domains like cyber-intelligence, law-enforcement sector, etc.

Liu et al. [16] proposed a modified fuzzy approach for cyber hate classification. This model uses four datasets collected from Twitter regarding four types of hate speech. Because the intersectionality of different types of hate speech is not addressed, more diversified features are extracted for hate speech detection. Fang et al. [17] proposed a multi-strategy sentiment analysis method with semantic fuzziness to extract the customer’s opinions expressed in sentiment Chinese phrases. The input reviews are taken from search forum website. Since the emotional phrases are not included in this work, calculation errors arise which results in lower accuracy.

A. Problem Statement

In the field of SA, machine learning methods such as Decision Tree [18], Logistic Regression [19], Support Vector Machine [20], Naive Bayes [21] and others have shown promising results, but most of them overly rely on hand-crafted features and necessitate a lot of manual design and adjustment, which is time-consuming and costly. Deep learning methods have achieved excellent performance with the help of large-scale corpus in many research fields, and become a research hotspot in SA [22]. However, in the majority of traditional text sentiment classification algorithms, the sentence-level sentiment classification remains difficult for various reasons, including a lack of semantic understanding and low classification accuracy. Moreover, the major existing methodology uses only 3 × 3 and 2D kernels which cannot be able to extract the review-text features effectively. Most of the current classification method commonly preprocesses only the basic stop-words, numbers, & symbols, and maintains the words which does not reveal any sentiment information (Ex: around, surround, door etc.). Keeping those words in the classification phase increases the complexity and time consumption of the method. In some existing literatures, when the length of a word in an input review is less than the length of the sentence matrix, the padding operation is used to pad with sufficient amount of NULL data in order to meet the original length of the sentence matrix. This padded NULL data reduces the impact of real data and dominates over it which results in less effective feature representation and negatively impacts the classification accuracy. Substantially, many classification methods classifies the reviews into two (happy, unhappy) or three classes (positive, negative, neutral) which results in an ineffective decision making process.

The main contribution of this paper is to propose a novel e-commerce review classification method namely RAVO based Multi-Size Kernel structured Text CNN based Review Classification (MSK-TCNN-RAVO) based on a new variant of CNN. This method utilizes a convolutional operation that employs multi-size, multi-dimensional kernels to generate enriched features which yields high review classification accuracy. This research demonstrates the concept of varying kernels applied on 1D kernels in order to extract features exclusively from accurate data rather than padded data, which is one of the key processes of this work. This work introduces a new concept called review-stop-words removal, in which, the words that do not convey any sentiment information are eliminated. Keeping these review-stop-words in classification increase the complexity and the time-consumption of both training and testing process. To our knowledge, there are no studies in the available literatures that address the removal of non-sentimental words during the classification stage. Subesquently, it is configured to generate five classification outcomes which will help companies make better business decisions. Finally, to improve the performance of the classifier, the proposed work is driven by a Rider driven African vulture optimization algorithm based on the overtaker's strategy of the Rider optimization algorithm.

III. Methodology

The proposed review-classifier network is designed based on Convolutional Neural Network (CNN). The design of this new MSK-TCNN-RAVO method is one-of-a-kind, allowing for correct classification of review data. The following are the unique aspects of the proposed work:

- Input processing for 7 × 7 size sentence matrix rather than a traditional 5 × 5 size.
- Design of multi-size Convolutional kernels.
- Integration of both 1D and 2D kernels during convolution process.
- Introducing the concept of varying kernel based convolution process to extract features exclusively from accurate data rather than padded data.
- Pooling layer constructed by dual pooling operations
- Configured to generate five classification results viz. Bad, Fair, Neutral, Good, and Excellent.

The aforementioned modifications make this MSK-TCNN-RAVO method as a novel work and more effective than existing Text-based CNN classifiers to classify the review-data. The proposed work is mainly divided into three sections such as: MSK-TCNN-RAVO review training process, MSK-TCNN-RAVO review testing process, and Optimization using Modified African Vulture Algorithm.

The workflow of the proposed work is shown in Fig. 1. The input reviews from three domains such as Laptop, Camera, Mobile is taken from Amazon website. The dataset as a whole has a lot of long sentences, with an average of 11.44 sentences per review. The proposed method has both training and testing sections. The basic stopwords, numbers, and symbols are removed from the input reviews in the preprocessing section and words which do not reveal any sentiment information is removed in review-stop-words removal section. The sentence matrix (SM) is generated by extracting each word from the enhanced reviews and placing it in matrix form. The above mentioned process is common for both training and testing section. Each review from the SM is fed as input to the proposed algorithm's training process and trained network is created. Similarly, in the testing section, each review from SM is fed as input to the proposed algorithm's testing process and it continues until all the reviews are processed. Finally, the review classification report is produced from the testing section.

A. MSK-TCNN-RAVO Review Training Process

1) Pre-processing: Reviews are usually composed of incomplete expression, a variety of noise and poorly structured sentences. Noise and unstructured Twitter data will affect the performance of tweet sentiment classification [23], [24]. Prior to feature selection, a series of preprocessing steps are performed on reviews to reduce the meaningless data in the review sentences.

Stop-words are words which have no value (positive or negative) in a sentiment analysis system that are meaningless in information retrieval, hence, they must be eliminated from the data set. For example stop words include "the", "as", "of", "and", "or", "to", etc. Stop word removing is substantial in the preprocessing, it has some advantages like reducing the size of stored data set and it improves the overall efficiency and effectiveness of the analysis system [25]. The proposed system uses a list of stop words obtained from Onix Text Retrieval Toolkit website [26].

In this proposed research work, three crucial pre-processing steps are carried out using text processing functions of MATLAB research tool to remove the (i) basic stop-words (ii) numbers, and (iii) symbols. The resultant pre-processed review is stored in the review-arrayR_{pp}.

2) Review-stop-words removal: The input of this section is the pre-processed review data. In the actual pre-processing section, the basic stop-words, numbers, and symbols are removed. But the remaining enhanced review contains the word which does not reveal any sentiment information, and those words are called as “Review-Stop-words”. Keeping those words will increase time consumption and the complexity of the classification process. A tool is developed to assist for the generation of review-stop-words. Any sample review dataset can be fed as input to this generalized tool which comprises two steps, viz. basic stop-words removal and unique words extraction. Afterwards, the review-stop-words are extracted manually and the review-stop-word list can be generated.

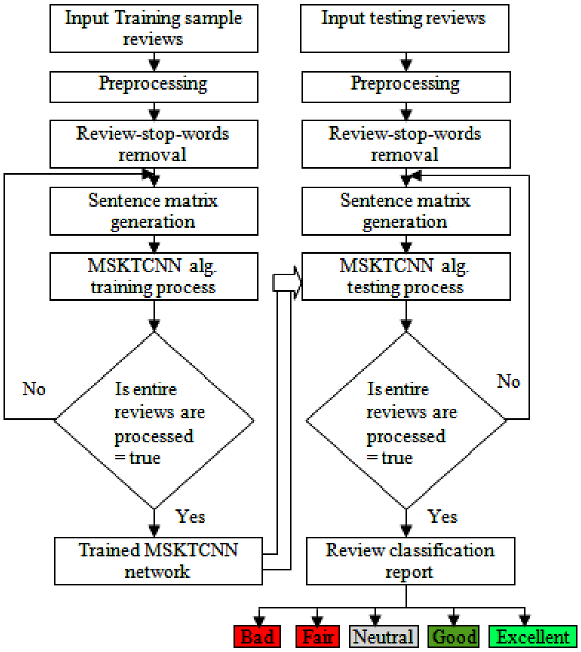

Input Training sample reviews
Preprocessing
Review-stop-words removal
Sentence matrix generation
MSKTCNN alg. training process
Is entire reviews are processed = true
No
Yes
Trained MSKTCNN network
Input testing reviews
Preprocessing
Review-stop-words removal
Sentence matrix generation
MSKTCNN alg. testing process
Is entire reviews are processed = true
No
Yes
Review classification report
Bad
Fair
Neutral
Good
Excellent

Fig. 1. Workflow of the Proposed MSK-TCNN-RAVO Classifcation Process.

SI No	Stop Words
1	today
2	seems
3	kind
4	something
5	after
6	around
7	across

Fig. 2. Sample List of Review-Stop-Words.

The sample review-stop-words are shown in Fig. 2.The survey-stop-words are loaded in the survey-stop-words-arrayL_{ssw}. In the pre-processed review, the survey-stop-words are removed based on the L_{ssw} list. The resultant survey after this removal process is stored in the review-array R_{ssw}.

3) Sentence matrix generation: The Sentence Matrix SM represents a sentence in the matrix form. The dimension of SM is fixed as 7×7. It is fixed by maintaining a trade-off between complexity and accuracy. A sentence of the enhanced review (after removing basic stop-words, numbers, symbols, and review-stop-words) R_{ssw} is extracted and each word of it is placed in the matrix form.

In this research work, the convolution process is carried out by taking SM as input using the combination of both 2D kernels such as 3×7 , 5×7 and 1D kernel such as 1×3, 1×5, and 1×7 in order to extract the review text feature representation effectively.

4) Training of MSK-TCNN-RAVO Algorithm: This algorithm is comprised of multi-size, multi-dimensional kernels. The varying kernel adopted in this technique retrieves features only from the exact data and not from the padded data. The dual pooling operations applied in this algorithm increase the efficiency of the feature map processing. The training process is carried out in order to produce five classification results. The architecture of the algorithm is depicted in Fig. 3.

a) 3×7 *Size Kernel based Convolution*: This convolution process is performed using six 3×7 size kernels such as k_0,k_1,k_2,k_3,k_4, and $\mathrm{k}_5$which is represented in Fig. 4. The column width of these kernels is fixed as seven, because the length of Sentence Matrix SM is seven. First the kernel k_0 is used to convolute the sentence matrix SM to produce convoluted matrix $\mathrm{CM}_{\mathrm{k0}}$. This process is performed using (1) to (5).

$$k_0 = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ -1 & 1 & 1 & 0 & 1 & 1 & -1 \\ -1 & -1 & -1 & 0 & -1 & -1 & -1 \end{bmatrix} \quad (1)$$

$$A = \sum_{m=0}^{3-1} \sum_{n=0}^{7-1} SM(i+m, j+n) * k_0(m,n), i = 0, j = 0 \quad (2)$$

$$B = \sum_{m=0}^{3-1} \sum_{n=0}^{7-1} SM(i+m, j+n) * k_0(m,n), i = 2, j = 0 \quad (3)$$

$$C = \sum_{m=0}^{3-1} \sum_{n=0}^{7-1} SM(i+m, j+n) * k_0(m,n), i = 4, j = 0 \quad (4)$$

$$CM_{k0} = \begin{bmatrix} A \\ B \\ C \end{bmatrix} \quad (5)$$

Herein, A, B, C represents the convoluted result by applying convolution on row of 0, 2, and 4 respectively inSM, and CM_{k0} indicates convoluted matrix by k_0 kernel.

Equation (1) defines the kernel k_0. Equation (2) computes the convoluted value by projecting the k_0 over the element of $SM(0,0)$. It means the convolution process performed on 0th row and 0th column in SM, and stored in A. Likewise the Equation (3) computes the convoluted result by considering the position $SM(2,0)$ and the output is assigned in B. The Equation (4) computes the convoluted result by considering the position $SM(4,0)$. Finally, the convoluted matrix CM_{k0} with size 3×1 is constructed by placing the A, B, and C values in the order using (5). Similarly the other 3×7 size kernels such as k_1, k_2, k_3, k_4, and k_5 is convoluted to find the convoluted matrix CM_{k1} , CM_{k2} , CM_{k3} , CM_{k4} , and CM_{k5} respectively.

b) 5×7 *Size Kernel based Convolution*: Here, the convolution process is performed using four 5×7 size kernels such as k_6 , k_7 , k_8 , and k_9 to generate four convoluted matrices such as CM_{k6}, CM_{k7}, CM_{k8}, and CM_{k9}. The four 5×7 size kernels are depicted in Fig. 5.

The kernel k_6 is used to compute the convoluted matrix CM_{k6} with size 6×1 using (6) and (7).

$$k_6 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 0 & 0 & 0 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (6)$$

$$\mathrm{CM}_{\mathrm{k6}} = \begin{bmatrix} \sum_{m=0}^{5-1} \sum_{n=0}^{7-1} SM(i+m, j+n) * k_6(m,n), i = 0, j = 0 \\ \sum_{m=0}^{5-1} \sum_{n=0}^{7-1} SM(i+m, j+n) * k_6(m,n), i = 1, j = 0 \\ \sum_{m=0}^{5-1} \sum_{n=0}^{7-1} SM(i+m, j+n) * k_6(m,n), i = 2, j = 0 \\ \sum_{m=0}^{5-1} \sum_{n=0}^{7-1} SM(i+m, j+n) * k_6(m,n), i = 3, j = 0 \\ \sum_{m=0}^{5-1} \sum_{n=0}^{7-1} SM(i+m, j+n) * k_6(m,n), i = 4, j = 0 \\ \sum_{m=0}^{5-1} \sum_{n=0}^{7-1} SM(i+m, j+n) * k_6(m,n), i = 5, j = 0 \end{bmatrix} \quad (7)$$

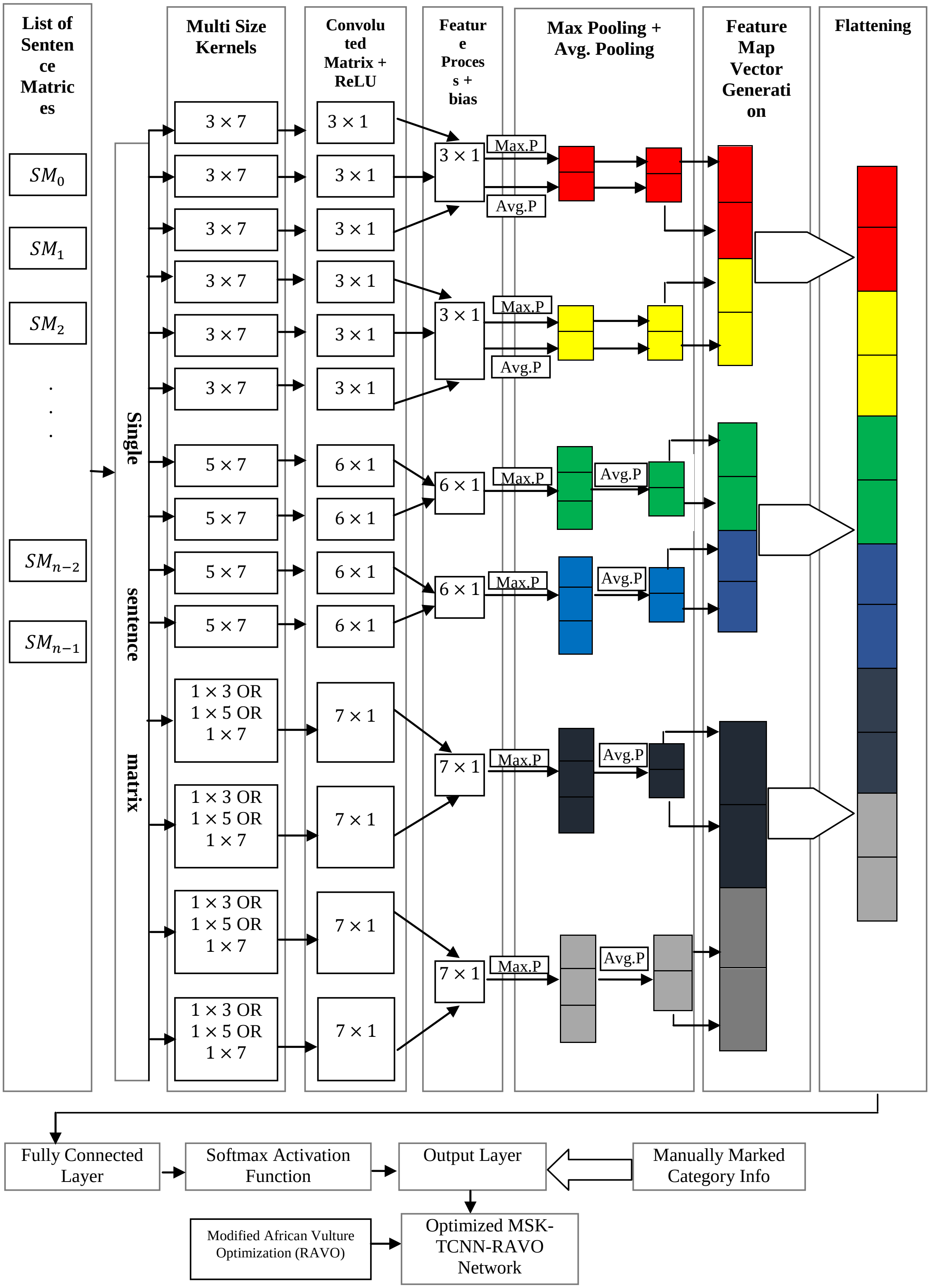
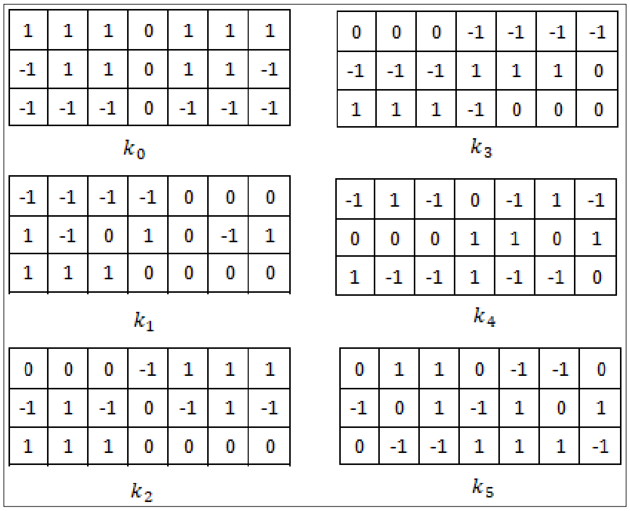


Fig. 3. Architecture Diagram of the Training Section of the Proposed MSK-TCNN-RAVO Method.



1	1	1	0	1	1	1
-1	1	1	0	1	1	-1
-1	-1	-1	0	-1	-1	-1

k_0

-1	-1	-1	-1	0	0	0
1	-1	0	1	0	-1	1
1	1	1	0	0	0	0

k_1

0	0	0	-1	1	1	1
-1	1	-1	0	-1	1	-1
1	1	1	0	0	0	0

k_2

0	0	0	-1	-1	-1	-1
-1	-1	-1	1	1	1	0
1	1	1	-1	0	0	0

k_3

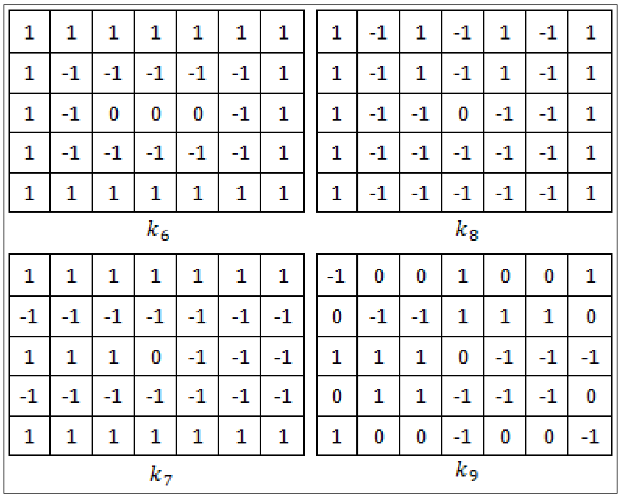
-1	1	-1	0	-1	1	-1
0	0	0	1	1	0	1
1	-1	-1	1	-1	-1	0

k_4

0	1	1	0	-1	-1	0
-1	0	1	-1	1	0	1
0	-1	-1	1	1	1	-1

k_5

Fig. 4. Representation of Six Kernel of 3 × 7 Size.



1	1	1	1	1	1	1
1	-1	-1	-1	-1	-1	1
1	-1	0	0	0	-1	1
1	-1	-1	-1	-1	-1	1
1	1	1	1	1	1	1

k_6

1	-1	1	-1	1	-1	1
1	-1	1	-1	1	-1	1
1	-1	-1	0	-1	-1	1
1	-1	-1	-1	-1	-1	1
1	-1	-1	-1	-1	-1	1

k_8

1	1	1	1	1	1	1
-1	-1	-1	-1	-1	-1	-1
1	1	1	0	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1
1	1	1	1	1	1	1

k_7

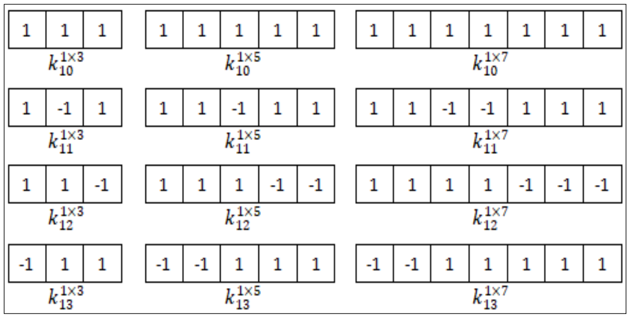
-1	0	0	1	0	0	1
0	-1	-1	1	1	1	0
1	1	1	0	-1	-1	-1
0	1	1	-1	-1	-1	0
1	0	0	-1	0	0	-1

k_9

Fig. 5. Representation of 5 × 7 Size Kernelsk_6, k_7,k_8, and k_9.

In Equation (7), the kernel k_6 is projected over the rows of padded-SM from row = 0 to row = 5, and the convolution process is performed. Similarly the other convoluted matrix such as CM_{k7}, CM_{k8}, and CM_{k9} is convoluted.

c) One Dimensional Varying Kernel based Convolution: In Sentence matrix SM, the length of a word may be less than the matrix length 7. In this case, padding operation pads the necessary quantity of NULL data to meet the length of the SM. However, the padded data diminishes the effect of real data in feature representation, lowering review classification accuracy. Hence, this research employs the concept of 1D varying kernels such as $1 \times 3, 1 \times 5$, 1×7 which have the length bounded in the range of 1×7 as shown in Fig. 6. The words which have length less than 7 in SM undergone the convolution operation with the aid of either the varying kernel 1×3 or 1×5. The words which have length equal to 7 in SM undergone the convolution process by using the kernel of 1×7 size. This concept extracts the features from the exact data and not from the padded data, which is one of the key processes of this research.



1	1	1

$k_{10}^{1\times3}$

1	1	1	1	1

$k_{10}^{1\times5}$

1	1	1	1	1	1	1

$k_{10}^{1\times7}$

1	-1	1

$k_{11}^{1\times3}$

1	1	-1	1	1

$k_{11}^{1\times5}$

1	1	-1	-1	1	1	1

$k_{11}^{1\times7}$

1	1	-1

$k_{12}^{1\times3}$

1	1	1	-1	-1

$k_{12}^{1\times5}$

1	1	1	1	-1	-1	-1

$k_{12}^{1\times7}$

-1	1	1

$k_{13}^{1\times3}$

-1	-1	1	1	1

$k_{13}^{1\times5}$

-1	-1	1	1	1	1	1

$k_{13}^{1\times7}$

Fig. 6. Illustration of Varying Kernels.

The convoluted matrix computation for the kernel k_{10} is performed using (8) to (14).

$$k_{10}^{1\times3} = \begin{bmatrix}1 & 1 & 1\end{bmatrix} \tag{8}$$

$$k_{10}^{1\times5} = \begin{bmatrix}1 & 1 & 1 & 1 & 1\end{bmatrix} \tag{9}$$

$$k_{10}^{1\times7} = \begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1\end{bmatrix} \tag{10}$$

$$H_{SM} = \text{HeightOfSentenceMatrix(SM)} \tag{11}$$

$$L_i = \text{Word Length } (SM_i)$$

$$i \in [0, H_{SM} - 1] \tag{12}$$

$$C_i = \begin{cases} \sum_{n=0}^{3-1} SM(i,n) * k_{10}^{1\times3}(0,n), if\ L_i <= 3 \\ \sum_{n=0}^{5-1} SM(i,n) * k_{10}^{1\times5}(0,n), else\ if\ L_i <= 5 \\ \sum_{n=0}^{7-1} SM(i,n) * k_{10}^{1\times7}(0,n), else\ if\ L_i <= 7 \end{cases}$$

$$i \in [0, H_{SM} - 1] \tag{13}$$

$$\mathrm{CM}_{\mathrm{k10}} = \begin{bmatrix} \mathrm{C}_0 \\ \mathrm{C}_1 \\ \mathrm{C}_2 \\ \mathrm{C}_3 \\ \mathrm{C}_4 \\ \mathrm{C}_5 \\ \mathrm{C}_6 \end{bmatrix} \tag{14}$$

where

H_{SM} - Height of the Sentence Matrix SM

Height Of Sentence Matrix – Function to find the height of sentence matrix SM

L_i - Length of ith word in SM

Word Length – Function to compute length of the specific word in SM

C_i - Convoluted value of 10th varying kernel corresponding to ith row of Sentence Matrix SM

n – Column movement index

$\mathrm{CM}_{\mathrm{k10}}$ - Convoluted matrix related to k_{10} kernel.

Equation (8) to Equation (10) defines the 10th varying kernel's variation such as $k_{10}^{1\times3}$, $k_{10}^{1\times5}$ and $k_{10}^{1\times7}$. Equation (11) computes the height of the Sentence Matrix SM and stores in the term H_{SM}. Equation (12) computes the length of each word of SM. Equation (13) computes the convoluted value of each word of SM using 10th varying kernel, by using the kernels $k_{10}^{1\times3}$ or $k_{10}^{1\times5}$ or $k_{10}^{1\times7}$ depending on the length of word L_i. The Equation (14) shows the convoluted matrix by merging the seven values of C_i. Similarly, the convoluted matrix such asCM_{k11}, CM_{k12}, and CM_{k13}will be computed.

d) ReLU Operation and Feature Matrix Generation: In this research, the ReLU function is used as activation function after convolution process. The function returns 0 if it receives any negative input, but for any positive value x, it returns the same value [27]. The ReLU activation function is designed for the convolution matrixCM_{k10} is based on Equation (15).

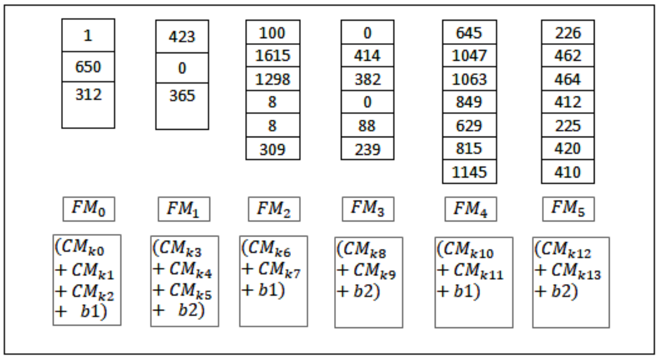
$$CM_{k0}^{i} = \left[\begin{array}{l} CM_{k0}^{i}, if CM_{k0}^{i} >= 0 \\ \quad 0, \qquad else \end{array}\right. \tag{15}$$

where $i \in [0, n-1]$

n – Height of the convoluted matrix CM_{k0} (Let it be 3 for 3×7 size kernels). Similarly the other convolution matrixes ranging CM_{k1} to $CM_{k13}$are computed based on (15).

The feature matrix is generated in order to reduce the quantity of convoluted matrix from 14 to 6. This helps to improve the speed of the review-analysis. Herein, the bias value participates its part similar with the traditional neural network. The bias value $b1$ is set as 1 and bias value $b2$ is set as 0. The illustration of feature matrix generation on the Sentence Matrix is shown in Fig. 7.

e) Pooling Operations: The pooling operation reduces the size of the feature map by keeping the key information. This research applies two types of pooling operations such as maximum pooling and Average pooling which essentially extracts the essence of the feature map. The feature matrix FM_0, FM_1 with 3×1 dimension is undergone the Max pooling and Avg. pooling operation which results a individual single value by computing the maximum value and average value respectively. The feature matrix FM_2, FM_3 with 6×1 dimension is undergone both the pooling operation based on Fig. 8.



FM_0	FM_1	FM_2	FM_3	FM_4	FM_5
1	423	100	0	645	226
650	0	1615	414	1047	462
312	365	1298	382	1063	464
		8	0	849	412
		8	88	629	225
		309	239	815	420
				1145	410
$(CM_{k0} + CM_{k1} + CM_{k2} + b1)$	$(CM_{k3} + CM_{k4} + CM_{k5} + b2)$	$(CM_{k6} + CM_{k7} + b1)$	$(CM_{k8} + CM_{k9} + b2)$	$(CM_{k10} + CM_{k11} + b1)$	$(CM_{k12} + CM_{k13} + b2)$

Fig. 7. Illustration of Feature Matrix Generation on the Example Sentence Matrix.

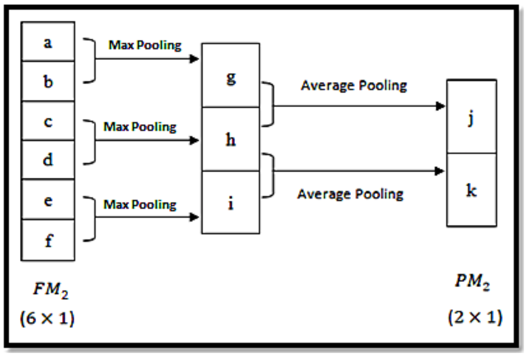


Fig. 8. Pooling Operation Performed in the Feature Map FM_2.

Similarly the feature matrix FM_4, FM_5 is undergone both the pooling operations as shown in Fig. 9. The illustration of pooling layer computation is depicted in Fig. 10. The 2×1 dimension is the resultant matrix after the pooling operations.

f) Feature Map Vector Generation: The feature map vector generation is performed to collect the same features of same size kernels into small groups. The size reduced 3×7 kernels are grouped in order to generate feature map vector FV_0 using (16).

$$FV_0 = \{ PM_{0,}^{0} PM_{0,}^{1} PM_{1,}^{0} PM_{1}^{1} \} \tag{16}$$

Similarly, the size reduced 5×7, and one dimensional kernel oriented pooled feature maps are grouped to generate the feature map vector FV_1, FV_2 respectively using (17) and (18).

$$FV_1 = \{ PM_{2,}^{0} PM_{2,}^{1} PM_{3,}^{0} PM_{3}^{1} \} \tag{17}$$

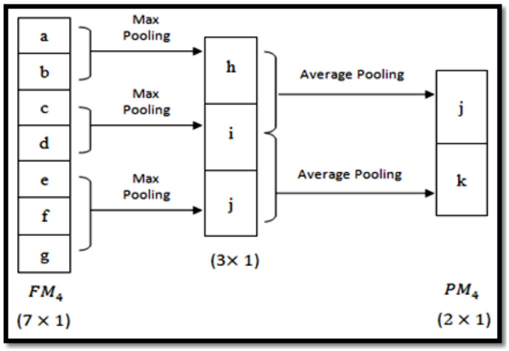
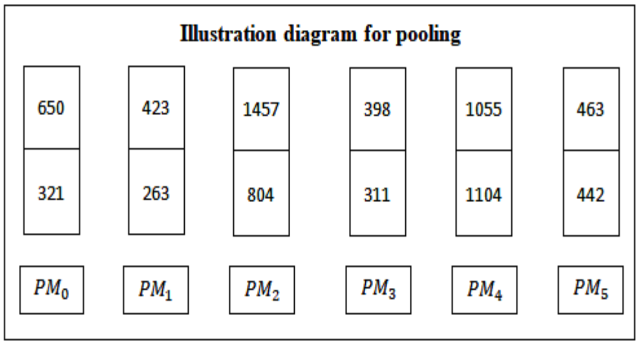


Fig. 9. Pooling Operation Performed in the Feature Map FM_4.



PM_0	PM_1	PM_2	PM_3	PM_4	PM_5
650	423	1457	398	1055	463
321	263	804	311	1104	442

Fig. 10. Illustration of Pooling Layer Computation.

Likewise, the feature map vector FV_2 is originated using the one dimensional kernels oriented pooled feature maps such as PM_4 and PM_5 based on (18).

$$FV_2 = \{ PM^0_{4,} PM^1_{4,} PM^0_{5,} PM^1_5 \} \tag{18}$$

g) Softmax Function: The resultant feature map vectors FV_0 , FV_1 , and FV_2 holds four elements each. The output of the convolution layer is flattened into a 1D array called as F, which is given as input to the fully connected layer. The output layer in an artificial neural network produces the given outputs for the program. In the training section of the proposed RAVO-MSKTCNN algorithm, the manually marked category information such as BD, FR, NL, GD, EX with target values 1, 2, 3, 4 and 5 respectively are fed to support the learning process. The extracted flattened features of the given training review sample is converted into probabilistic distributions form by softmax function. This data is compiled by the target values of training categories until it converges by the back propagation concept. Thus the training of a review data is progressed. Multi reviews are trained using the MSK-TCNN-RAVO algorithm, and the trained network of the same is generated.

B. MSK-TCNN-RAVO Review Testing Process

The testing process of proposed algorithm is used to classify the review data with the aid of the trained network. The Fig. 11 depicts the architecture of the testing process of the proposed MSK-TCNN-RAVO network. The basic stop words, numbers, and symbols are removed from the test reviews in the preprocessing section and words which do not reveal any sentiment information are removed in review-stop-words removal section.

The enhanced reviews SM is fed as input to the testing section of the MSK-TCNN-RAVO algorithm and the convolution operations such as 3×7 , 5×7 , and one dimensional varying kernel based convolution are performed on SM to generate the corresponding convoluted matrices. The ReLU function is used as an activation function after convolution process, and feature matrix is generated in order to reduce the quantity of convolution matrices from 14 to 6 as shown in Fig. 7. The dual pooling operation such as Max and Average pooling is applied and finally the feature map vector is generated. The extracted feature map is given to the MSK-TCNN-RAVO network as a feature which is then classified into five classes as shown in Fig. 11. This process is repeated until all the test reviews are processed. The classification report generated by the proposed algorithm can be used for better decision making process.

Algorithm -1 of review testing process

```
Step 1: The test review is given as input
Step 2: Remove basic stop words from the given test review
Step 3: Remove the survey-stop-words from the pre-processed test reviews
Step 4: Convert the processed test review into a Sentence Matrix (SM) of size $7 \times 7$
Step 5: Design six $3 \times 7$ size kernels such as $k_0$, $k_1$,$k_2$,$k_3$,$k_4$, and $k_5$.
Step 6: Convolute the Sentence Matrix (SM) by $k_0$ kernel and find the Convolution Matrix $CM_{k0}$
Step 7: Repeat Step 6 with the other kernels such as $k_1$,$k_2$,$k_3$,$k_4$, and $k_5$ to obtain the Convolution Matrix $CM_{k1}$, $CM_{k2}$, $CM_{k3}$, $CM_{k4}$, and $CM_{k5}$ respectively.
Step 8: Design four $5 \times 7$ size kernels such as $k_6$, $k_7$,$k_8$,and $k_9$.
Step 9: Convolute the Sentence Matrix (SM) by $k_6$ kernel and find the Convolution Matrix $CM_{k6}$.
Step 10: Repeat Step 9 with the other kernels such as $k_7$,$k_8$, and $k_9$ to obtain the Convolution Matrix $CM_{k7}$ , $CM_{k8}$ , and $CM_{k9}$ respectively.
Step 11: Construct four one dimensional varying kernels of size $1 \times 3$ , $1 \times 5$ , $1 \times 7$ . Name the four kernels as $k_{10}$ , $k_{11}$,$k_{12}$,and $k_{13}$.
Step 12: Compute Convolution matrix $CM_{k10}$ via one dimensional varying kernels $k_{10}^{1\times3}$, $k_{10}^{1\times5}$, and $k_{10}^{1\times7}$.
Step 13: Compute convoluted matrix $CM_{k11}$ using kernels $k_{11}^{1\times3}$, $k_{11}^{1\times5}$, and $k_{11}^{1\times7}$
Step 14: Compute Convolutional matrix $CM_{k12}$ using kernels like $k_{12}^{1\times3}$, $k_{12}^{1\times5}$ and $k_{12}^{1\times7}$
Step 15: Compute convoluted matrix $CM_{k13}$ using kernels such as $k_{13}^{1\times3}$, $k_{13}^{1\times5}$, and $k_{13}^{1\times7}$.
Step 16: Activate the convoluted matrices such as $\mathrm{CM}_{\mathrm{k0}}$ to $\mathrm{CM}_{\mathrm{k13}}$using the ReLU activation function.
Step 17: Construct the feature matrix $FM_0$ to $FM_5$ using the convoluted matrices $CM_{k0}$to $CM_{k13}$.
Step 18: Apply the max pooling and average pooling operations on the feature matrix $FM_0$ to $FM_5$ and compute the corresponding pooled matrix $PM_0$to $PM_5$ of $2 \times 1$ size.
Step 19: Construct $3 \times 7$ kernel based feature map vector $FV_0$ to $FV_2$.
Step 20: Feed the output of fully connected layer to Softmax activation function.
Step 21: Output classification results for test reviews based on RAVO driven MK-TCNN network.
```

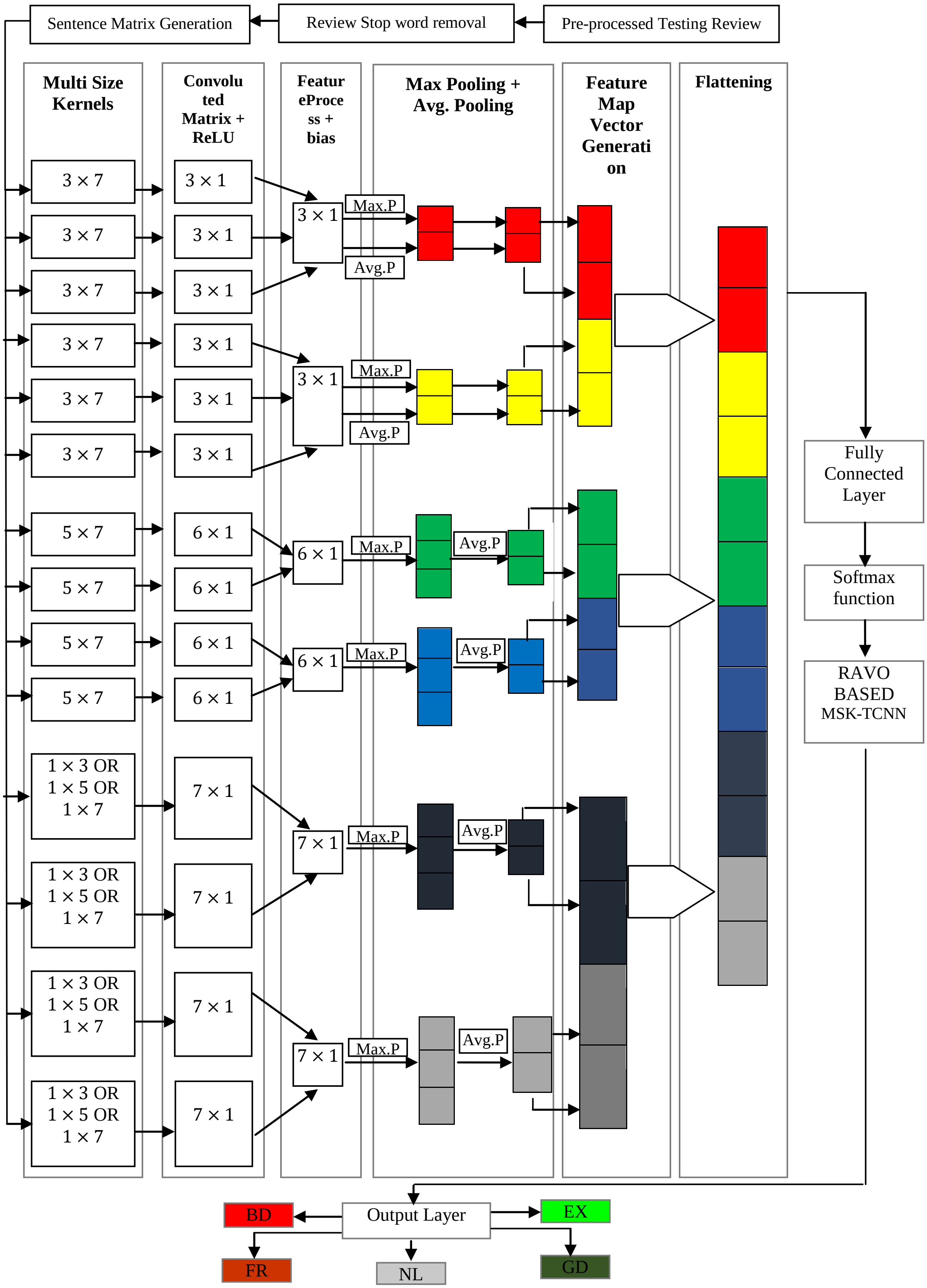


Fig. 11. Architecture Diagram of the Testing Section of the Proposed MSK-TCNN-RAVO Method.

C. RAVO Driven MSK-TCNN

The proposed MSK-TCNN-RAVO network is driven by Rider driven African vulture optimization algorithm based on over taker's strategy of rider optimization algorithm. The parameters of the MSK-TCNN-RAVO is tuned by considering the significant parameters of RAVO such as multi kernel size (M_{ks}), Pooling type (P_o), and Feature map Vector (F_{mv}). These parameters are analyzed using (19).

$$T_p = \{M_{ks=1\ to\ t} \,||F_{mv=1\ to\ t} \,||P_{o=1\ to\ t}\} \tag{19}$$

Where, T_p denotes the hidden layers to be optimized in which t represents the initialization parameters of RAVO from MSK-TCNN-RAVO. In addition to the hidden layers, number of epoch, learning rate, batch size are considered as the hyper parameters for obtaining maximum efficiency in terms of performance matrices such as accuracy, precision, recall, F-Score, Time taken, MSE, Misclassification ratio, ranking index, etc. Moreover, these parameters are taken from MSK-TCNN to evaluate the performance based on Mean Square Error (MSE) and Logarithmic Loss. The MSE is considered as a fitness value of MSK-TCNN-RAVO that is processed until it reaches the minimal MSE. Equation (20) is considered as fitness evaluation.

$$F_{(T_p)} = \min(MSE\ , lss) \tag{20}$$

After the initialization of the fitness value the following equation is utilized for determining the best solution.

$$R_i = \begin{cases} Best_{sol1} & if\ p_i = L_1 \\ Best_{sol2} & if\ p_i = L_2 \end{cases} \tag{21}$$

The selection of best solution is obtained by Roulette wheel selection strategy as shown in Equation (22).

$$p_i = \frac{F_{(T_p)}}{\sum_{iter=1}^{n} F_{(T_p)}} \tag{22}$$

Where, R_i and $F_{(T_p)}$ are, respectively, the probability solution and the fitness solutions of each individual as i. This process is repeated until reaches the n number of iterations. This strategy enhances the diversity in AVOA. Then the second strategy of AVOA is rate of starvation which is determined using (23) and (24).

$$s = t \times \left(sin^w\left(\frac{\pi}{2} \times \frac{iter}{maxiters}\right) + \cos\left(\frac{iter}{maxiter}\right) - 1\right) \tag{23}$$

$$V_s = (2 \times rand_1 + 1) \times z \times \left(1 - \frac{iter}{maxiter}\right) + s \tag{24}$$

Where, V_s denotes the satiated vultures, i denotes the iteration, $maxiter$ denotes the maximum iteration, z denotes the random number in range (-1 and 1), t represents the random value in range (0 and 1). Here, the starvation of vulture is determined by $z < 0$ and $z > 0$ conditions, that show vulture is starved and not starved, respectively. According to this RAVO is executed exploration strategy.

$$\text{P (i+1)} = \begin{cases} \mathrm{R(i) - D(i)} \times V_s \text{ if } \mathrm{P_1} \geq \mathrm{rand_{P_1}} \\ \mathrm{X \times R(i) - P(i)} \text{ if } \mathrm{P_1} < \mathrm{rand_{P_2}} \end{cases} \tag{25}$$

Where, $P(i+1)$ is new position vector, V_s denotes the number of vulture gratified using (24) in the current iteration. Moreover, the variables in the following denotes

X➔ Coefficient vector to maximize the random motion that is measured using $\mathrm{X} = 2 \times \mathrm{rand}$ in which rand is random number in range of 0 and 1.

P (i)➔ The position vectors of the current iteration.

$$\mathrm{P(i+1) = R(i)} - V_s + \mathrm{rand_2} \times ((\mathrm{up_b} - \mathrm{lw_b}) \times \mathrm{rand_3} + \mathrm{lw_b}) \tag{26}$$

In aforementioned Equation (26), $\mathrm{rand_2}$ is the random value in range of 0 and 1. $\mathrm{lw_b}$, and $\mathrm{up_b}$ are the variable that lower bound and upper bound, respectively. The Equation (26) is utilized to generate random solution based on the range of $\mathrm{lw_b}$, and $\mathrm{up_b}$. The "$\mathrm{rand_3}$", increases the coefficient of random. Here, $\mathrm{rand_3}$ reaches nearly 1, it disseminates the solution with similar order that increases random motion based $\mathrm{lw_b}$. Therefore, the diversity can be increased and explored maximum search space solution. Then, the model is moving on exploitation strategy if the $|F_{(T_p)}|$ has obtained below 1.

D. Exploitation

The exploitation has incorporated with two strategies where $\mathrm{P_2}$ and $\mathrm{P_3}$ are the parameters to determine the strategy to be executed. The ranges of these parameters are initialized at 0 and 1 before executing the searching process.

(i) Phase 1

In the first phase, the exploitation is obtained only when the value of $\left|F_{(T_p)}\right|$ reaches between the range of 1 and 0.5. Here, determining the $\mathrm{P_2}$ value is highly significant before the execution of searching process with the range between 0 and 1. Here, the "$\mathrm{rand_{P2}}$", generates random number between 0 and 1. If $\mathrm{rand_{p2}} \geq \mathrm{P_2}$, the Siege-fight strategy is executed. Conversely, if $\mathrm{rand_{p2}} < \mathrm{P_2}$, rotation flight is processed. The mathematical formula of this is shown in (27).

$$\mathrm{P(i+1)} = \begin{cases} \mathrm{D(i)} \times \left(F_{(T_p)} + \mathrm{rand_4}\right) - \mathrm{d(t)}\text{if } \mathrm{P_2} \geq \mathrm{rand_{p2}} \\ \quad \mathrm{d(t) = R(i) - P(i)} \qquad \text{if } \mathrm{P_2} \geq \mathrm{rand_{p2}} \end{cases} \tag{27}$$

1) Competition for food: If the condition has obtained $\left|F_{(T_p)}\right| \geq 0.5$, means that the vultures in the environment are highly satiated and they have considerable energy for surviving. This can be determined by Equation (28).

$$D(i) \times (V_s + rand) - d(t)$$

$$d(t) = R(i) - P(i) \tag{28}$$

Where, $D(i)$ is evaluated using (27), V_s denotes the rate of vulture satiated, $rand_4$ denotes the random number ($0\ and\ 1$) helps for generating random coefficient.

Then the vulture nest strategy is rotating flight that is based on spiraling between whole vulture and one of the two best vultures. The mathematically formulation of these strategy is shown in (29) and (30).

$$S_1 = R(i) \times \left(\frac{rand_5 \times p(i)}{2\pi}\right) \times \cos(p(i)) \tag{29}$$

$$S_2 = R(i) \times \left(\frac{rand_6 \times p(i)}{2\pi}\right) \times \sin(p(i)) \tag{30}$$

$$P(i+1) = R(i) - (S_1 + S_2) \tag{31}$$

Where,

$R(i)$ is executed based on same strategy as earlier mentioned. Cos and Sin are utilized in this work based on sine and cosine function. $rand_5$ and $rand_6$ are random number generated in the range of 0 and 1. S_1, and S_2 are generated using (29) and (30). Equation (31) to update the newly generated solution.

(ii) Phase 2:

In the second phase of exploitation, the two vultures' movements accumulate several types of vultures over the food source, and the siege and aggressive strive to find food are carried out. If $\left|F_{(T_p)}\right| \leq 0.5$, this phase is relying on $rand_{p3}$ in the range (0 and 1). If $rand_{p3} \geq p_3$, various vultures are accumulating over the food. Otherwise, if $rand_{p3} < p_3$, the aggressive siege flight strategy is executed. The mathematical formulation of this is shown in equation (32).

$$P(i+1) = \begin{cases} equ\ (31) if\ rand_{p3} \geq P_3 \\ equ\ (32) if\ rand_{p3} < P_3 \end{cases}$$

$$A_1 = Best_{Vult_1}(i) - \frac{Best_{Vult_1}(i) \times p(i)}{Best_{Vult_1}(i) \times p(i)^2} \times F$$

$$A_2 = Best_{Vult_2}(i) - \frac{Best_{Vult_2}(i) \times p(i)}{Best_{Vult_2}(i) \times p(i)^2} \times F$$

$$(i+1) = \frac{A_1 + A_2}{2} \tag{32}$$

a) Aggressive Competition for Food

If $|F_{(T_p)}|$ < 0.5 occurs, repositioning the vultures using equation (33)

$$P(i+1) = R(i) - |d(t)| \times over_t\ (d) \tag{33}$$

Here, the updating strategy of Rider Optimization is executed especially based on over taker's strategy. This strategy helps to enhance the performance of AVOA. Therefore, Rider based enhanced AVOA can make highly effective performance.

$$over_{t+1}^{0}(\text{iter, k}) = over_t(i,k) + [d^Z(t)] * X^{hv}(hv,k)$$

Where, $over_t(i,k)$, represents the iter_{th} vulture in the k_{th} coordinates, $d^Z(t)$, represents the direction of i_{th} rider at the time t, and in the range (-1 and 1) $X^{hv}(hv,k)$, represents the head vulture in the k_{th} coordinates. Moreover, the computation of $d^Z(t)$ is achieved by using (34) to (36).

$$d^Z(t) = \left[\frac{2}{1-\log(S_t^R(iter))}\right] - 1 \tag{34}$$

$$S_t^R(i) = \frac{r_t(i)}{max_{i=2}^{R} r_t(i)} \tag{35}$$

$$r_t\ (i) = \frac{1}{||X_i - L_T||} \tag{36}$$

$S_t^R(i)$ represents the relative success rate of the i_{th} vulture at the time t and it will be in the interval of 0 and 1. Finally, determine the coordinate selection based on normalized distance vector that obtained by gauging the difference of i_{th} vulture and head vulture using equation (37).

$$l(i,j) = |X_t(i,j) - X^h(h,j)| \tag{37}$$

The value of j is selected for $l(i,j)th$ values are less than its fitness value. Therefore, the Rider's over taking strategy improves the performance by replacing Levy fight strategy. The RAVO can provide high performance than traditional African Vulture Optimization algorithm.

Pseudo Code: RAVO driven MSKTCNN

```
Start:
Define the parameters of the RAVO-MSKTCNN
I/P
    Five sets of Review datasets are divided into 70% for Training and 30% for Testing
O/P
    Optimized RAVO-MSKTCNN
    Minimized MSE, & Cross Entropy and Maximized accuracy, precision, recall, Fscore
Initialize the parameters of RAVO-MSKTCNN
    RAVO Parameter
        Number of population, Number of iteration,
    MSKTCNN Parameters:
        Batch Size, Number of Epoch, Learning Rate, Multi-Size Kernel, Pooling Size
Determine the MSKTCNN parameter using equation (19)
    $T_p = \{M_{ks=1\ to\ t} || F_{mv=1\ to\ t} || P_{o=1\ to\ t}\}$
While (($MSE$ & $Loss$) ≤ $termination\ Criteria$ ) do
    Best solutions are obtained using $R_i$
    For (each solution ($p_i$)) do
        Equation (21) to obtain $R_i$
    Update the solution using (23) to (35)
Rider's Overstake's strategy to improve AVOA using
        (34) to (37)
Return
Optimized Parameters of RAVO based MSKTCNN obtained
End
```

IV. Analysis and Discussion

This section is very important in any research paper which portrays the efficiency of the proposed or the other existing methods. The proposed MSK-TCNN-RAVO classification task is compared with three existing methods mentioned below:

1) Weakly supervised deep embedding for review Sentence sentiment classification (WS_RSSC) [28].

2) Sentiment similarity analysis based mining of user's trust from e-commerce reviews (SSA_TER) [29].

3) CNN based multiple convolution and pooling for text sentiment classification (CNN_MCP_TSC) [30].

In order to test the proposed MSK-TCNN-RAVO classifier, the training and testing reviews are taken from five benchmark databases. The product reviews from different categories such as Mobile, Laptop, and Camera are taken from Amazon website. The names of the different datasets are given below:

1) Dataset of Laptop Product Review (DS1_LPR) [28].

2) Dataset of Camera Product Review (DS2_ CPR) [28].

3) Dataset of Mobile Product Review for NOKIA 3310 (DS3_MPR3310) [28].

4) Dataset of Mobile Product Review for SAMSUNG EPIC 4G (DS4_MPRSE4G) [31].

5) Dataset of Mobile Product Review for NOKIA 105 (DS5_MPR105) [32].

From each dataset, 7000 reviews are extracted, out of that, 4000 reviews are utilized for training and the remaining 3000 reviews which are completely new to the MSK-TCNN-RAVO classifier is used for testing purpose to ensure the classifier's efficiency. Following that, 2000 more reviews are extracted from the training samples to act as test reviews, totaling 5000 test reviews for each dataset.

A. *Classification Accuracy*

The Accuracy of the classifier is computed using (38) and its values are represented in Table I.

$$Accuracy = \frac{TP+TN}{(TP+FP+FN+TN)} \tag{38}$$

A true positive (TP) is an outcome where the model correctly predicts the positive class. Similarly, a true negative (TN) is an outcome where the model correctly predicts the negative class. A false positive (FP) is an outcome where the model incorrectly predicts the positive class. And a false negative (FN) is an outcome where the model incorrectly predicts the negative class.

TABLE I. CLASSIFICATION ACCURACY ANALYSIS FOR FIVE DATABASES AND FOUR METHODS

Database Name	**Classification Accuracy (%)**				
	WS_RSSC Method	**SSA_TER Method**	**CNN_MCP_TSC Method**	**Proposed MSK-TCNN-RAVO Method (Before RAVO)**	**Proposed MSK-TCNN-RAVO Method (After RAVO)**
DS1_ LPR	78.2	79	80.5	**86.7**	**90.6%**
DS2_ CPR	77.6	77.3	78.9	**84**	**89.6%**
DS3_ MPR3310	80.1	81.5	82.4	**89.6**	**93.2%**
DS4_ MPRSE4G	81.8	82.6	84	**90.5**	**95.5%**
DS5_ MPR105	79.3	80.4	81.4	**88.9**	**92.4%**

The accuracy of the proposed model is high in all datasets when compared to other models. The highest accuracy is produced for the DS4_MPRSE4G dataset which is 95.5% after performing optimization via RAVO. The result shows that the proportion of classified true results produced by the proposed work is high when compared to other models.

B. *Precision*

Precision measures the exactness of a classifier. It can be calculated using (39).

$$Precision = \frac{TP}{TP+FP} \tag{39}$$

The average precision values are represented in Table II. The precision is computed for each method for all the five datasets. Average precision of a specific method is computed by averaging the precision value of five datasets. The precision value for a specific dataset is computed by testing 5000 test review samples. The proposed methodology has a higher precision value, indicating that the proposed model produces less false positives, implying that the proposed model's prediction is more reliable. The average value of the proposed method is 0.89

TABLE II. AVERAGE PRECISION ANALYSIS FOR FOUR METHODS

Method	**Average Precision**
WS_RSSC Method	0.74
SSA_TER Method	0.80
CNN_MCP_TSC Method	0.79
Proposed MSK-TCNN-RAVO Method	**0.89**

C. *Recall*

Recall measures the completeness, or sensitivity, of a classifier. Higher recall means less false negatives, while lower recall means more false negatives. The Recall can be calculated using (40).

$$Recall = \frac{TP}{(TP+FN)} \tag{40}$$

The average recall values are depicted in Fig. 12. The MSK-TCNN-RAVO method achieves the higher recall value of 0.895 which indicates that the percentage of true positives correctly classified by the proposed model is high compared to other existing models.

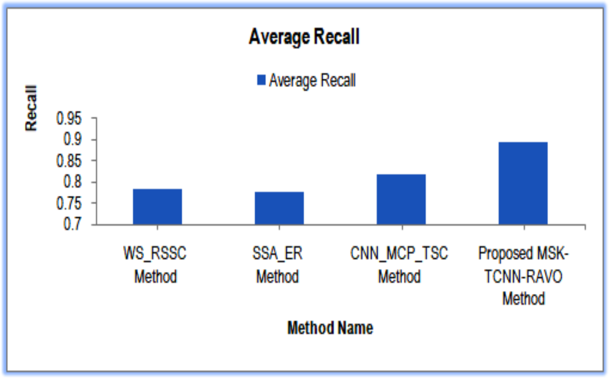


Fig. 12. Average Recall Analysis for Four Methods.

D. F-Score

In F-Score, precision and recall can be combined to produce a single metric known as F-measure, which is the weighted harmonic mean of precision and recall. The F-score can be calculated using (41).

$$Fscore = \frac{2*(precision*Recall)}{(Precision+Recall)} \tag{41}$$

The F-Score analysis is represented in Fig. 13. The F-Score value corresponding to a specific method and particular database is computed by processing the 5000 test review samples. Highest F-score value 0.947 is achieved by the MSK-TCNN-RAVO method for the DS4_MPRSE4G dataset.

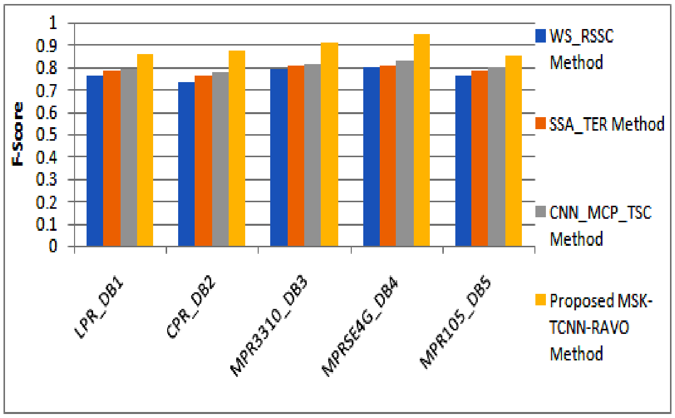


Fig. 13. F-Score Analysis for Five Database and Four Methods.

E. Mean Square Error

In Machine Learning, main goal is to minimize the error which is defined by the Loss Function. It can be computed using (42). The MSE analysis values are listed in Table III.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(PR_i - GT_i)^{\,2} \tag{42}$$

Here, N denotes the No. of samples tested, PR_iindicates the predicted class for ith sample, and GT_i denotes the ground truth class for ith sample.

It can be observed from the Table III that the MSE value for the proposed method is drastically minimized after the RAVO optimization. The least error value after optimization is 426, which indicates that the proposed method's prediction error rate is very low when compared to the other three current models. The WS_RSSC method produces the highest error value of 1090 in this analysis for the DS1_LPR dataset.

F. Cross Entropy

This metric is used in neural networks when a classifier's output is multiclass prediction probabilities. In general, minimizing categorical cross-entropy gives greater accuracy for the classifier [33]. It can be computed using (43), and it's computed values are depicted in Fig. 14.

$$LogarithmicLoss\ (lls) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\ * \log(p_{ij}) \tag{43}$$

TABLE III. MSE ANALYSIS FOR FIVE DATABASES AND FOUR METHODS

Database Name	MSE				
	WS_RSSC Method	**SSA_TER Method**	**CNN_MCP_TSC Method**	**Proposed MSK-TCNN-RAVO Method (Before RAVO)**	**Proposed MSK-TCNN-RAVO Method (After RAVO)**
DS1_ LPR	**1090**	1050	975	**665**	**615**
DS2_ CPR	1120	1135	1055	**800**	**745**
DS3_ MPR3310	995	925	880	**520**	**473**
DS4_ MPRSE4G	910	870	800	**475**	**426**
DS5_ MPR105	1035	980	930	**555**	**503**

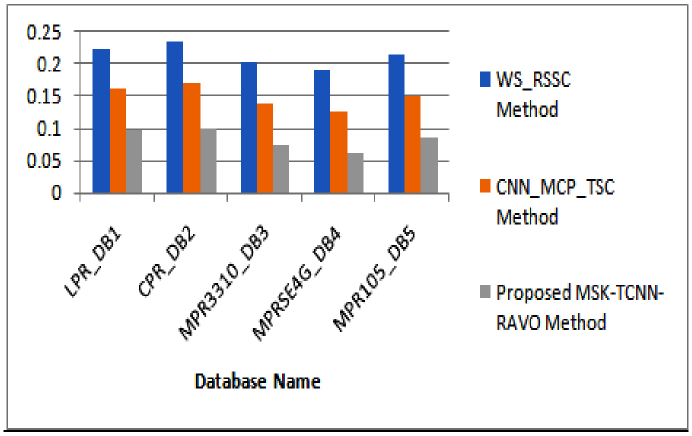


Fig. 14. Cross Entropy Analysis for Five Databases and Three Methods.

Here,$y_{ij} = 1$, if the sample i belongs to class j else 0. p_{ij} is the probability of the classifier that predicts of sample i belonging to class j. Since, the SSA_TER method is not coded based on neural network, this analysis is performed for the other three methods and all datasets. The proposed method gives a least value of **0.**0619 after RAVO optimization on DS4_MPRSE4G dataset which means it gives high probability for classified data with more stabilized value.

G. Time Taken

In time taken analysis, the time consumed by the methods to produce the classification results is evaluated. The time taken by all methods is listed in Table IV.

In this analysis, WS_RSSC method takes the least execution time for all datasets. But, when considering the classification accuracy of the proposed method, the excess time taken by the proposed model is acceptable. The time consumption difference between WS_RSSC and the MSK-TCNN-RAVO classifier is very less i.e., 1.26 seconds. Similarly, the time consumption difference between CNN_MCP_TSC and the proposed classifier is only 0.76 seconds.

TABLE IV. TIME TAKEN ANALYSIS FOR FIVE DATABASES AND FOUR METHODS

Database Name	Time Taken (in seconds)			
	WS_RSSC Method	SSA_TER Method	CNN_MCP_TSC Method	Proposed MSK-TCNN-RAVO Method
DS1_ LPR	5.472	5.892	7.513	**6.726**
DS2_ CPR	5.593	5.986	7.742	**6.995**
DS3_ MPR3310	5.401	5.802	7.298	**6.497**
DS4_ MPRSE4G	5.512	5.923	7.632	**6.883**
DS5_ MPR105	5.456	5.841	7.396	**6.654**

H. *Misclassification Ratio*

This analysis is carried out to determine the ratio of incorrect classification results generated by the classifier. It can be computed using (44). The ratios of misclassified results generated by all the methods are depicted in Fig. 15.

$$Misclassification = \frac{FP+FN}{Total\ Data} \qquad (44)$$

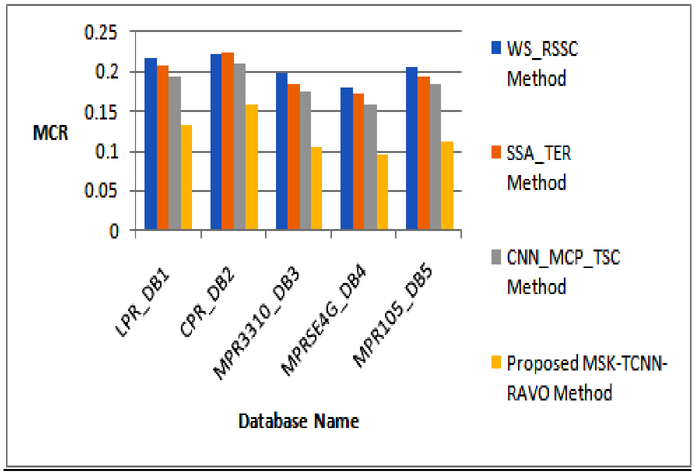


Fig. 15. Misclassification Ratio Analysis for Five Databases and Four Methods.

From the analysis, the misclassification ratio produced by the proposed method is very low for DS4_MPRSE4G dataset which is 0.095. It shows that the proposed classifier produces very low incorrect classification results. The second best method which has less misclassification ratio is CNN_MCP_TSC which has 0.160 for the same dataset.

I. *Performance based Ranking İndex*

This analysis ranks the review classification methods based on the ranking index. The ranked values are depicted in Fig. 16.

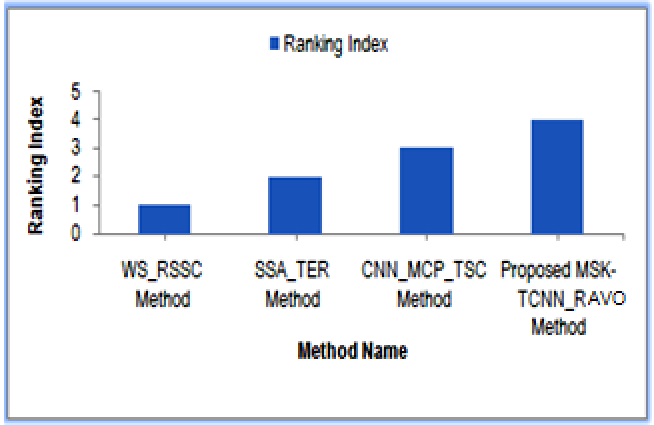


Fig. 16. Ranking İndex Analysis for the Four Methods.

According to classification accuracy, precision, f-score, recall, mean square error, cross entropy, misclassification results, and the ranking index of each method is decided. The best method among the four methods is set with rank 4, i.e., the highest rank. The second best method is ranked by 3 and so on. Hence, from the above Fig. 16, the proposed method achieves the highest fourth rank which shows it is the best method for review classification than the other existing methods.

V. CONCLUSION

In this work, a classification method namely MSK-TCNN-RAVO is proposed for e-commerce reviews. The proposed algorithm is executed on five different datasets from three different domains utilized for experimentation and efficient analysis, with 7000 review samples per database being used for training and testing the proposed classifier. The MSK-TCNN-RAVO classifier achieves the highest classification accuracy of 95.5% when compared to other three existing algorithms such as WS_SSC, SSA_TER, and CNN_MCP_TSC which achieves the accuracy of 81.8%, 82.6%, and 84% respectively. The highest F-Score value of 0.947 is achieved by the proposed method when compared to other three methods. The proposed MSK-TCNN-RAVO classifier outperforms the other three approaches in all metrics, and it consistently performs well on the DS4_MPRSEG4 dataset. When compared to the WS_RSSC technique, the proposed classifier improves classification accuracy by 8.54% before RAVO and 12.86% after RAVO optimization. Furthermore, the MSK-TCNN-RAVO classifier achieves the highest values on all evaluation metrics such as Classification accuracy, Precision, Recall, F-Score, MSE, Cross entropy, and Misclassification ratio, which proves that the proposed method is well suited for opinion mining. When processing the emoticons in the input reviews, the computing complexity of the proposed work grows. Additionally, a slight tweak to the method will be planned to address the increased CPU utilization observed when the emojis are being trained into the classifier. In future work, it is intended to modify the coding to extend the MSK-TCNN-RAVO classifier to produce hepta classifier which classifies the reviews into seven mode classification results. Furthermore, in order to further enhance the classification accuracy and reduce processing time, an extended stop words list will be employed as a future improvement to eliminate basic stop words and survey stop words in the preprocessing phase.

DECLARATION

Competing Interests

- There are no conflicts of interest.

Funding

- There is no funding available.

Authors' contributions

- Methodology design, analysis, dataset selection, experiment evaluation is performed by the first author
- Correction is performed by the second author

ACKNOWLEDGMENT

I would like to express my heartfelt gratitude to our university for offering unwavering support in all aspects to complete our research project successfully.

REFERENCES

[1] "Wikipedia". [Online].Found at: https://en.wikipedia.org/wiki/Online_shopping [Accessed On: 26-November-2021].

[2] "IGIGlobal".[Online].Found at: https://www.igi-global.com/dictionary/using-the-flipped-classroom-to-improve-knowledge-creation-of-masters-level-students-in-engineering/21327 [Accessed On: 12-December-2021].

[3] A. Yousaf, "Emotion recognition by textual tweets classification using voting classifier (LR-SGD)", IEEE Access, 9, pp. 6286–6295, 2021.

[4] R. Ren and D. Wu, "An innovative sentiment analysis to measure herd behavior", IEEE Trans. Syst. Man, Cybern.Syst., 50(10), pp. 3841–3851, 2020.

[5] P. Sasikala and L. Mary Immaculate Sheela, "Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS", J. Big Data, 7(1), 2020.

[6] Z. Shaukat, A. A. Zulfiqar, C. Xiao, M. Azeem, and T. Mahmood, "Sentiment analysis on IMDB using lexicon and neural networks", SN Appl. Sci., 2(2), 2020.

[7] Z. Jia, X. Bai, and S. Pang, "Hierarchical gated deep memory network with position-aware for aspect-based sentiment analysis", IEEE Access, 8, pp. 136340–136347, 2020.

[8] Y. Cheng, L. Yao, G. Xiang, G. Zhang, T. Tang, and L. Zhong, "Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism", IEEE Access, 8, pp. 134964–134975, 2020.

[9] M. Dong, Y. Li, X. Tang, J. Xu, S. Bi, and Y. Cai, "Variable convolution and pooling convolutional neural network for text sentiment classification", IEEE Access, 8, pp. 16174–16186, 2020.

[10] V. Gupta, A. Aggarwal, and T. Chakraborty, "Detecting and characterizing extremist reviewer groups in online product reviews", IEEE Trans. Comput. Soc. Syst., 7(3), pp. 741–750, 2020.

[11] M. M. Madbouly, S. M. Darwish, and R. Essameldin, "Modified fuzzy sentiment analysis approach based on user ranking suitable for online social networks", IET Softw., 14(3), pp. 300–307, 2020.

[12] R. Bhalla and A. Bagga, "Opinion mining framework using proposed rb-bayes model for text classification", Int. J. Electr.Comput.Eng., 9(1), pp. 477–484, 2019.

[13] S. Zhang and H. Zhong, "Mining users trust from E-Commerce reviews based on sentiment similarity analysis", IEEE Access, 7, pp. 13523–13535, 2019.

[14] A. Abdul Aziz and A. Starkey, "Predicting supervise machine learning performances for sentiment analysis using contextual-based approaches", IEEE Access,8, pp. 17722–17733, 2020.

[15] F. Iqbal, "A Hybrid framework for Sentiment analysis using genetic algorithm based feature reduction", IEEE Access, 7, pp. 14637–14652, 2019.

[16] H. Liu, P. Burnap, W. Alorainy, and M. L. Williams, "A fuzzy approach to text classification with two-Stage training for ambiguous instances", IEEE Trans. Comput. Soc. Syst., 6(2), pp. 227–240, 2019.

[17] Y. Fang, H. Tan, and J. Zhang, "Multi-strategy sentiment analysis of consumer reviews based on semantic fuzziness", IEEE Access, 6, pp. 20625–20631, 2018.

[18] H. Zou, X. Tang, B. Xie, and B. Liu, "Sentiment classification using machine learning techniques with syntax features", Proc. Int. Conf. Comput. Sci. Comput. Intell.2015, 5(4), pp. 175–176, 2016.

[19] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis", ACL-HLT-Proc. 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol., 1, pp. 142–150, 2011.

[20] T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources", Proc. Conf. Empir.Methods Nat. Lang. Process.pp. 412–418, 2004.

[21] S. Tan, X. Cheng, Y. Wang, and H. Xu, "Adapting naive bayes to domain adaptation for sentiment analysis", Lect. Notes Comput. Sci, 5478, pp. 362–374, 2009.

[22] K. Zhang, M. Jiao, X. Chen, Z. Wang, B. Liu, and L. Liu, "SC-BiCapsNet: A Sentiment classification model based on bi-channel capsule network", IEEE Access, 7, pp. 171801–171813, 2019.

[23] J. Zhao, "Pre-processing boosting twitter sentiment analysis?", Proc. - IEEE Int. Conf. Smart City, SmartCity, pp. 748–753, 2015.

[24] Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on twitter sentiment analysis", IEEE Access, 5, pp. 2870–2879, 2017.

[25] V. Singh and B. Saini, "An effective tokenization algorithm for Information", pp. 109–119, 2014.

[26] "Onix Text Retrieval Toolkit API". [Online].Found at: http://www.lextek.com/manuals/onix/stopwords1.html [Accessed On: 30-October-2021].

[27] "DeepAI".[Online].Found at: https://deepai.org/machine-learning-glossary-and-terms/relu [Accessed On: 13-November-2021].

[28] W. Zhao., "Weakly-supervised deep embedding for product review sentiment analysis", IEEE Trans. Knowl.Data Eng., vol. 30(1), pp. 185–197, 2018.

[29] S. Zhang, H. Zhong, "Mining Users Trust From E-Commerce Reviews Based on Sentiment Similarity Analysis", IEEE Access, vol. 7, pp. 13523–13535, 2019.

[30] M. Dong, Y. Li, X. Tang, J. Xu, S. Bi, Y. Cai, "Variable Convolution and Pooling Convolutional Neural Network for Text Sentiment Classification", IEEE Access, vol. 8, pp. 16174–16186, 2020.

[31] Kaggle, "Amazon Reviews: Unlocked Mobile Phones", 2021. [Available: https://www.kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones. [Accessed On: 14-January-2021].

[32] Kaggle, "Amazon Mobile Reviews", 2021. [Available: https://www.kaggle.com/binaryjoker/amazon-mobile-reviews. [Accessed On: 26-February-2021].

[33] "TowardsDataScience". [Online].Found at: https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226 [Accessed On: 26-December-2021].

Comparison of Image Enhancement Algorithms for Improving the Visual Quality in Computer Vision Application

Jenita Subash[1]*
Department of ECE
Cambridge Institute of Technology, Bangalore, India

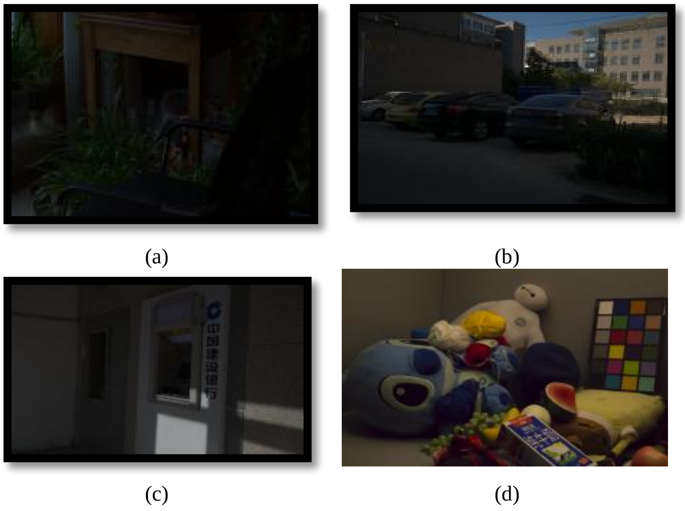
Dr. Jharna Majumdar[2]
Department of CSE
Cambridge Institute of Technology, Bangalore, India

***Abstract*—Computer vision has its numerous real-world applications on Visual Object Tracking which includes human-computer interaction, autonomous vehicles, robotics, motion-based recognition, video indexing, surveillance and security, human-computer interaction, autonomous vehicles, robotics, motion-based recognition, video indexing, surveillance and security. The factors affecting the tracking process is due to low illumination, haze and cloudy environment and noisy environment. In this paper, we aim to extensively review the latest trends and advances in adaptive enhancement algorithm and evaluate the performance using Full reference like, SSIM (Structure Similarity Index Measure), MS-SSIM (Multi-scale Structure Similarity Index Measure), ESSIM (Edge Strength Structural Similarity Index), FSIM (Feature Similarity Index Measure), VIF (Visual Information Fidelity), CW-SSIM (complex wavelet structural similarity), UQI (Universal Quality Index), IEF (Image Enhancement Factor), IQI (Image Quality Index), EME (Enhancement Measurement Error), CVSI (Contrast and Visual Salient Information), MCSD (Multiscale contrast similarity deviation), NQM (Noise Quality Measure), Gradient Magnitude Similarity Mean (GMSM), Gradient Magnitude Similarity Deviation (GMSM) and no-reference image quality measures Perception based Image Quality Evaluator (PIQE), Blind/Reference less Image Spatial Quality Evaluator (BRISQUE), Naturalness Image Quality Evaluator (NIQE), Average Gradient (AG), Contrast, Information Entropy (IE), Lightness order Error (LOE). The main purpose of adaptive image enhancement is to smooth the uniform area and sharpen the border of an image to improve its visual quality. In this paper, fourteen image enhancement algorithms were tested on LoL dataset to benchmark the time taken to process them and their output quality was evaluated. Results from this study will give insights to image analysts for selecting image enhancement algorithms which acts as a pre- processing stage for Visual object Tracking.**

***Keywords*—Tracking; robotics; surveillance; enhancement**

I. Introduction

In weakly illuminated environments, the images and video quality are often degraded. This leads to reduction in the performance of particular systems, such as those used in consumer electronics, visual surveillance and intelligent traffic analysis. For example, the low lighting conditions in nighttime environments can produce images and video with low contrast, which reduces the visibility [1]. Digital images used with contemporary imaging- and vision-related applications [2] and capturing the image in inappropriate lighting environment have a low-light effect, deficient contrast, and improper colors [3]. Therefore, it is very difficult to capture images with high-quality in that the low-light effect environment which may reduce the performance related to image processing and computer vision applications [4, 5], and such images usually comprise of vast dark regions with reduced visibility [6]. Samples of such images are shown in Fig. 1. There exist various image enhancement algorithms for improving the quality of images acquired under cloudy or other conditions.



(a)

(b)

(c)

(d)

Fig. 1. Various Types of Low-light Images. (a) Night Time Image; (b) Unevenly illuminated Image; c) Shadowed Environment Image; d) Image with a Dark Appearance.

Lowlight images are images that have a dark appearance, have uneven illumination, and they are captured in a shadowed environment [7]. The input parameters for these algorithms can vary from very minimal to relatively extensive. Depending on the algorithm the time taken to process an image can also vary. Based on the type of algorithm used and the input parameters specified the output quality of the resultant image will be different. Given with numerously available image enhancement algorithms, it is not feasible to evaluate all of them to determine their suitability. The primary objective of this study is to evaluate a suite of commonly used

*Corresponding Author.

image enhancement algorithms on low-illumination images. In the first phase of this study, fourteen image enhancement algorithms like Improved Type-II Fuzzy Set-based Algorithm, Retinex-based Multiphase Algorithm, Fusion-based enhancing method, Adaptive Image Enhancement Method for Correcting Low-Illumination Images, Fast efficient algorithm for enhancement of low lighting, A Multiscale Retinex, Bio-inspired multiexposure fusion frame work, Deep low light image enhancement, Adaptively Increasing Value Histogram Equalization (AIVHE)) were tested on LOL dataset to benchmark the quality of the output image and the time taken to process them.

Due to the rapid development of image enhancement technology, various enhancement algorithms such as retinex model [9–11], fuzzy theory [12, 13], Fusion based approach [14, 15], Deep learning Approach [16,17], Histogram Equalization based approach [18,19] etc. were developed. For example, as shown in Fig. 2, around 250 literatures on image enhancement algorithms were studied. The methods involved mainly include histogram equalization, Retinex model, Fusion based Approach, Fuzzy based approach and deep learning methods. Each of the image enhancement methods has their own advantages as well as disadvantages. The eye of a human has the ability of filtering the influence of light and obtains the reflectivity of the surface of the object to determine colour. Therefore, the formation of a low-light image can be described as follows:

$$L(x,y)=R(x,y)\cdot B(x,y) \tag{1}$$

where L(x, y) is the original image, R(x, y) is the reflection image, B(x, y) is the illuminance image and (x, y) is the pixel coordinates.

In this paper, we provide the progress of image enhancement algorithms during the past two decades. We mainly introduce the image enhancement methods separately in three aspects based on supervised methods, unsupervised methods and quality evaluation. The block diagram of the whole framework is shown in Fig. 3 in this paper.

The rest of paper is organized as follows. Section III introduces the image enhancement techniques based on Fuzzy based, Retinex based and Fusion based, Histogram based and Deep learning-based approach. Section IV elaborates in detail the image quality assessment using Full-reference, No-reference and Image Error Measurement. Section V deal with the results and discussions. Sections VI elaborates about results and discussions.

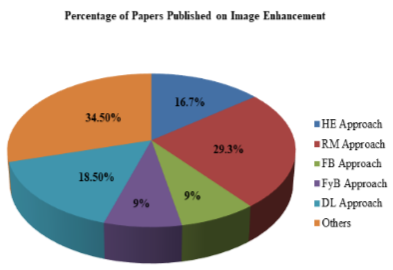


Fig. 2. Statistics of Percentage of Papers Published on Image Enhancement.

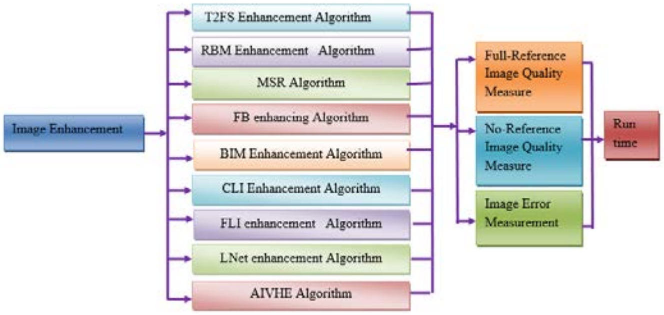


Image Enhancement
T2FS Enhancement Algorithm
RBM Enhancement Algorithm
MSR Algorithm
FB enhancing Algorithm
BIM Enhancement Algorithm
CLI Enhancement Algorithm
FLI enhancement Algorithm
LNet enhancement Algorithm
AIVHE Algorithm
Full-Reference Image Quality Measure
No-Reference Image Quality Measure
Image Error Measurement
Run time

Fig. 3. Block Diagram of the Workflow.

II. Datasets Used

The LOL dataset comprise of 500 low-light and normal-light image pairs and divided into 485 training pairs and 15 testing pairs. Most of the images are indoor scenes. The resolution of all images is 400×600.

III. T2FS, RB and HE based Image Enhancement

A. T2FS[20]

A Type-II fuzzy set (T2FS) based algorithm [20] was introduced for enhancing the contrast of grayscale medical images. This algorithm improves the contrast by Fuzzifying the image. Then, apply the Type-II fuzzy membership function are determined with the lower and upper ranges of the Hamacher t-conorm, where, α is a parameter that controls the amount of contrast enhancement, in that it should satisfy $0 < \alpha \leq 1$, when $\alpha > 0.6$, better contrast enhancement is obtained [20]. An improved type-II fuzzy set (IT2FS) algorithm [21], using Fuzzified image followed by Hamacher t-conorm method and then finally applying Gamma Correction The enhanced output of Improved Type-II fuzzy set-based algorithm with different α values is as shown in the Fig. 4(a) When α is between 0.3.5 and 0.55, the results will be obtained with satisfactory visual quality. When increasing α, the brightness is reduced while the contrast is enhanced selecting the proper value of α leads to desired results. To produce satisfactory results the proper gamma value can be around 0.50.

B. Retinex based Algorithm

Zou, Y et al. [22] and Kallel, F et al. [23] introduced various image enhancement algorithms for contrast enhancement in CT images. There exists different low-intricacy concept which improves the image illumination. Among such concepts, the single-scale retinex (SSR) model proposed by Jobson et al. [24] was examined because it involves simple calculations and improves the illumination of images. In brief, the SSR model works by estimating an illumination image from its degraded counterpart by performing a discrete convolution (*) between a degraded image and a discrete 2D Gaussian surround function (DGSF) [25].

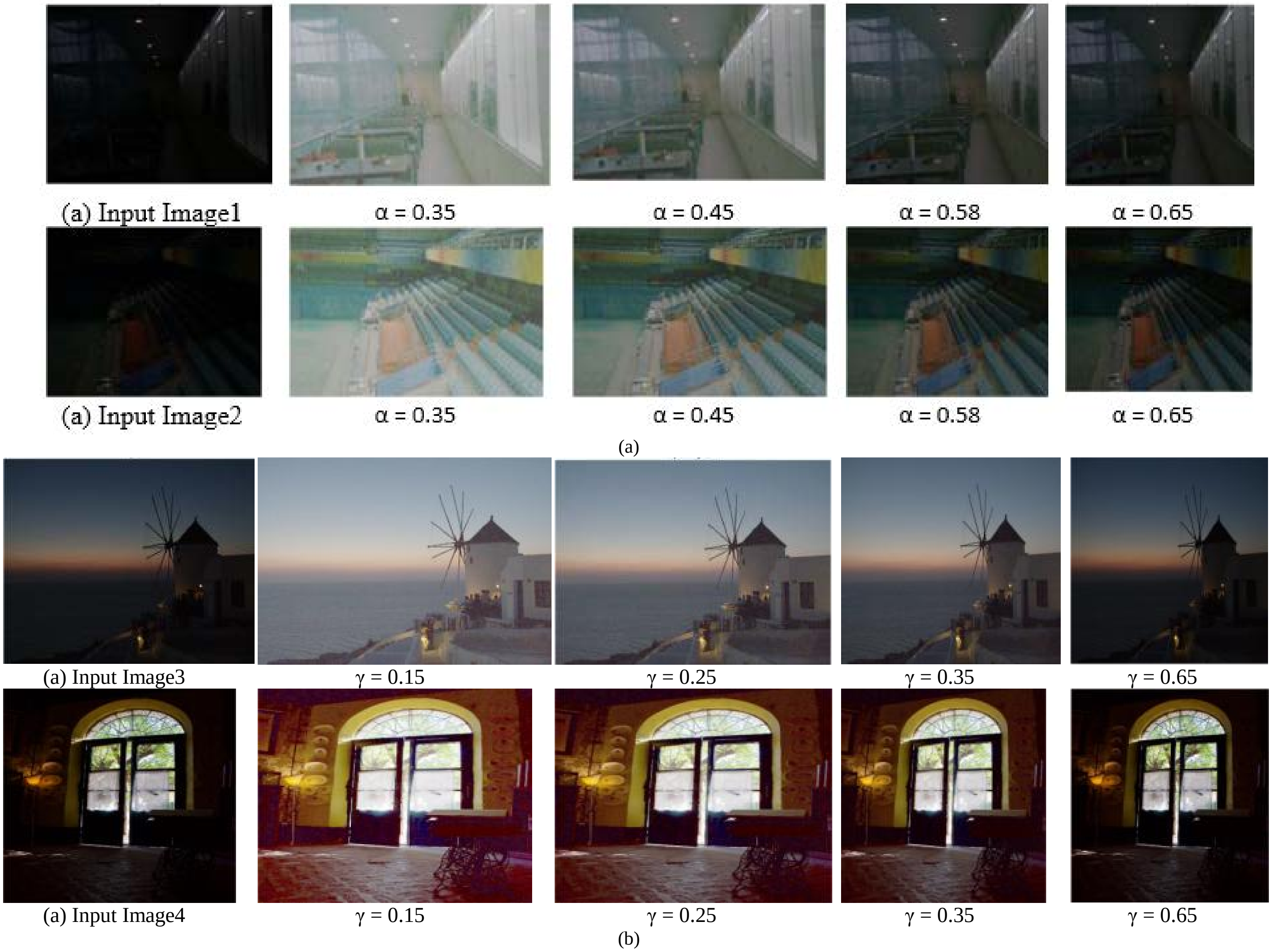


Fig. 4. (a) Type-II Fuzzy Set-based Algorithm with different α Values, (b) Retinex-based Multiphase Algorithm with different γ Values.

1) RBMA [26]: Mohammad Abid et al. [26] proposed RBMA which involves in determining the log of the illumination and the original images followed by computation of GCS (Gama Corrected Sigmoid function) .The enhanced output of Retinex-based Multiphase Algorithm with different values of γ is as shown in the Fig. 4(b).

Intensive experiments reveal that acceptable quality results are obtained when the γ value is between 0.1 and 0.35.

2) FBEM [27]: Xueyang Fu et al. [27] employed an illumination estimating algorithm based on morphological closing image and an illumination image. The two inputs - improved and contrast-enhanced versions of the first decomposed illumination were derived using the sigmoid function and adaptive histogram equalization. Designing two weights based on these inputs, an adjusted illumination is produced by fusing the derived inputs with the corresponding weights in a multi-scale fashion. Through a proper weighting and fusion strategy, the advantages of different techniques are blended to produce the adjusted illumination. The final enhanced image is obtained by compensating the adjusted illumination back to the reflectance.

In this fusion-based framework, images under different weak illumination conditions such as non-uniform illumination, backlighting, and nighttime can be enhanced.

3) AIEM [28]: Wencheng Wang et al. [28] proposed Adaptive Image Enhancement Method for Correcting Low-Illumination Images. The original RGB image is converted to HSV color space, and the V component is used to extract the illumination component of the scene using the multiscale Gaussian function. Then based on the Weber-Fechner law, a correction function is constructed, and two images are obtained through adaptive adjustments to the image enhancement function parameters based on the distribution profiles of the illumination components. Finally, an image fusion strategy is formulated and used to extract the details from the two images. Compared with the classic algorithm, the AIEM algorithm can improve the overall brightness and contrast of an image and the enhanced images appear clear, bright, and natural.

4) FEAE [29]: Xuan Dong et al. [29] proposed a Low lighting video enhancement algorithm by applying the invert operation on low lighting video frames, and then performing haze removal on the inverted video frames, before performing the invert operation again to obtain the output video frames.

5) LIME [30]: Xiaojie Guo et al. [30] proposed an effective low-light image enhancement (LIME) method. More concretely, the illumination of each pixel is first estimated individually by finding the maximum value in R, G and B channels. Further, we refine the initial illumination map by imposing a structure prior on it, as the final illumination map. Having the well-constructed illumination map, the enhancement can be achieved accordingly.

6) BIMEF [31]: Zhenqiang Ying et al. [31] proposed a framework mainly consists of four main components:

The first component, named Multi-Exposure Sampler, determines how many images are required and the exposure ratio of each image to be fused; the second component, named Multi-Exposure Generator, use a camera response model and the Specified exposure ratio to synthetic multi-exposure images; the third component, named Multi-Exposure Evaluator, determines the weight map of each image when fusing; the last component, named Multi-Exposure Combiner, is to fuse the generated images to the final enhanced result based on the weight maps.

7) SRIE [32]: In this paper, a weighted variational model for simultaneously estimating reflectance and illumination is presented. First, by analyzing the characteristic of the logarithmic transformation, we show that the logarithmic transformation is not proper to be directly used as regularization terms. Then, based on the previous analysis, a weighted variational model is introduced for better prior representation and an alternating minimization scheme is adopted to solve the proposed model.

8) NPEA [33]: Shuhang Wang et al. [33] proposed an enhancement algorithm for non-uniform illumination images. In general, this paper makes the following three major contributions. First, a lightness-order error measure is proposed to access naturalness preservation objectively. Second, a bright-pass filter is proposed to decompose an image into reflectance and illumination, which, respectively, determine the details and the naturalness of the image. Third, a bi-log transformation is applied, which is utilized to map the illumination to make a balance between details and naturalness.

9) BPHE [34]: In Brightness Preserving Bi-Histogram Equalization (BBHE) [34], the Input image is splitted into two sub images based on the mean of the input image. Samples of the input image which are less than or equal to mean forms one sub image, the other sub image consists of samples which are greater than the mean. Each of these sub images are independently equalized based on their respective histograms. The first sub image, containing samples less than or equal to mean, are mapped into the range from the minimum gray level to the input mean. The second sub image, containing samples greater than the mean are mapped into the range from the mean to the maximum gray level.

10)MSRA [36]: Daniel J. Jobson et al. [36] extend the designed single-scale center/surround retinex to a multiscale version that achieves simultaneous dynamic range compression/color consistency/ lightness rendition. This extension fails to produce good color rendition for a class of images that contain violations of the gray-world assumption implicit to the theoretical foundation of the retinex. Therefore, we define a method of color restoration that corrects for this deficiency at the cost of a modest dilution in color consistency.

11)LightenNet [38]: The purpose of LightenNet [38] is to learn a mapping, which takes a weakly illuminated image as input and outputs its illumination map that is subsequently used to obtain the enhanced image based on Retinex model. The architecture is LightenNet. LightenNet consists of four convolution layers, *i.e.*, patch extraction and representation, feature enhancement, nonlinear mapping, and reconstruction.

IV. Image Quality Assessment

Image Quality Assessment (IQA) is considered as a characteristic property of an image. Degradation of perceived images is measured by image quality assessment. Usually, degradation is calculated compared to an ideal image. Quality of image can be described technically as well as objectively to indicate the deviation from the ideal or reference model. It also relates to the subjective perception or prediction of an image [8], such as an image of a human look. Image Quality Assessment is grouped into two categories based on the availability of a reference image. The categories of Image Quality assessment methods are as shown in Fig. 5.

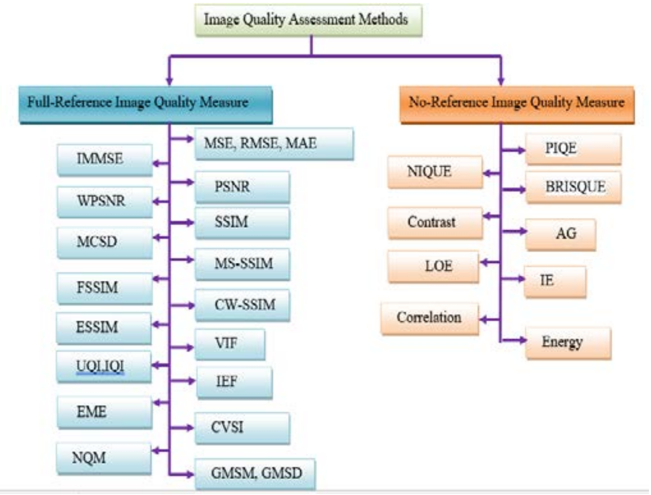


Fig. 5. Categories of Image Quality Assessment Methods.

V. Results and Discussion

A comparison is made with fourteen methods that are, T2FS [20], RBMA [26], FBEM [27], AIEM [28], FEAE [29], LIME [30], BIMEF [31], SRIE [32], NPEA [33], BPHE [34], CAVIEHE [35], MSRA [36], MSRCR [37], LightenNet [38] and the outcomes of such comparisons are evaluated by 30 metrics. Fig. 6 to 9 demonstrates the comparison results. Table I to Table XXIX exhibit the recorded metrics scores and processing times of the conducted comparison. Fig. 6 demonstrates the comparison results. Fig. 10 shows the GMS map for the entire different algorithm (Table I to Table XXIX) exhibit the recorded metrics scores and processing times of the conducted comparison.

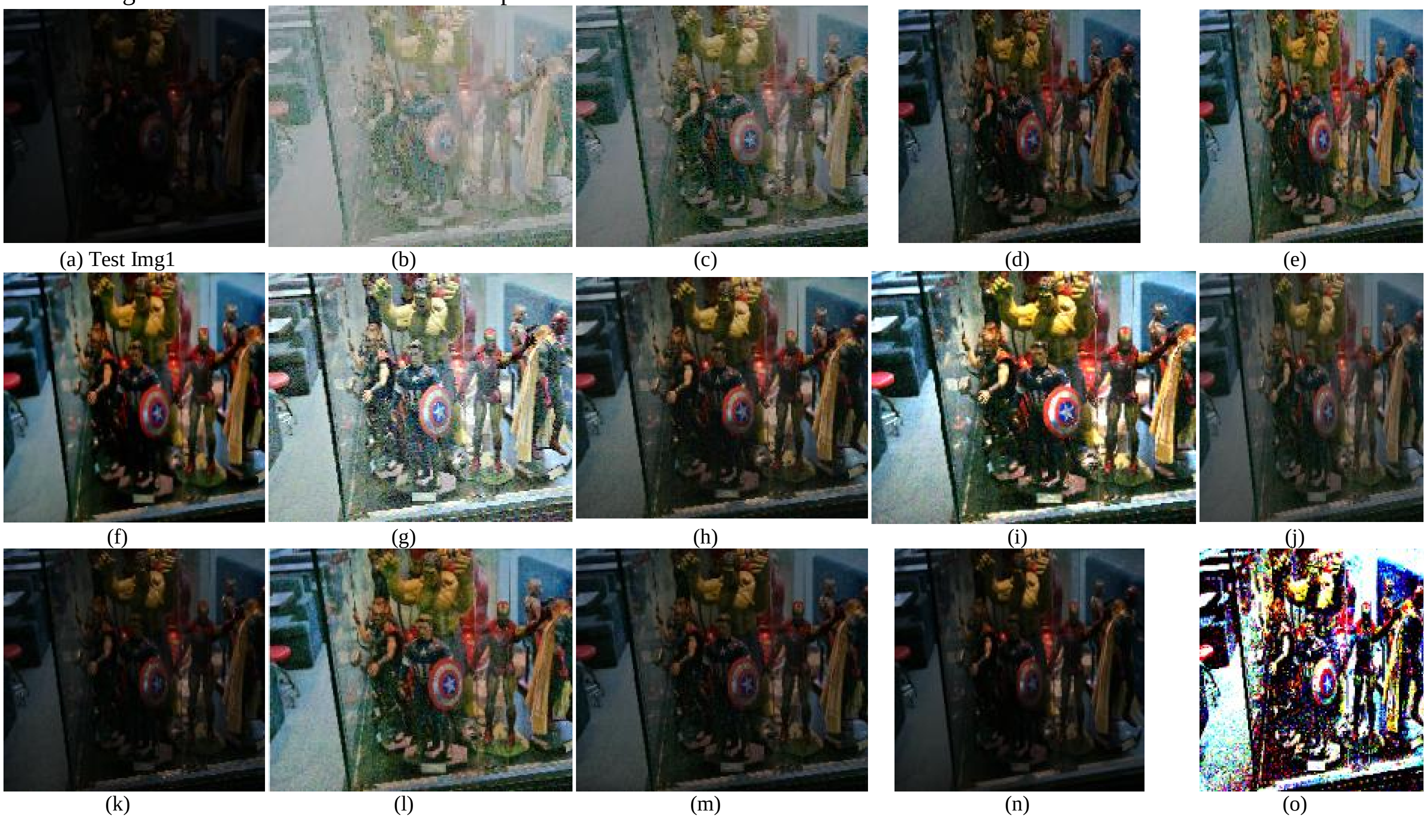

(a) Test Img1 (b) (c) (d) (e)
(f) (g) (h) (i) (j)
(k) (l) (m) (n) (o)

Fig. 6. The Comparison Outcomes Test Image1 (a) Real Low-light Image; The following Images are enhanced by: (b) IT2FB [20], (c) RBMA[26], (d) FBEM [27], (e) AIEM [28], (f) FEAE [29], (g) MSRA [30], (h) CAVIEHE [31], (i) LIME [32], (j) BIMEF [33], (k) LNET [34], (l) NPEA[35] [m] SRIE [36] [n]BPHE[37] [o] MSRCR [38].

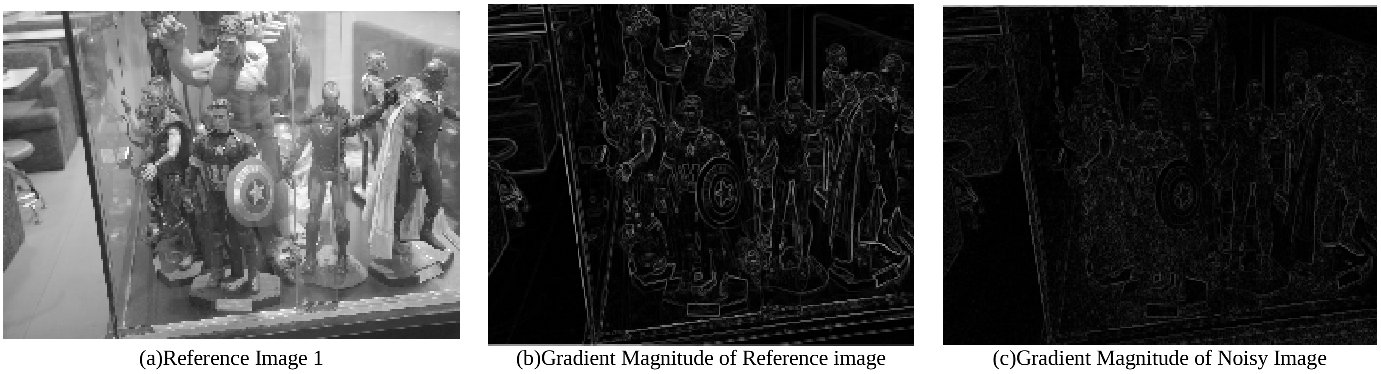
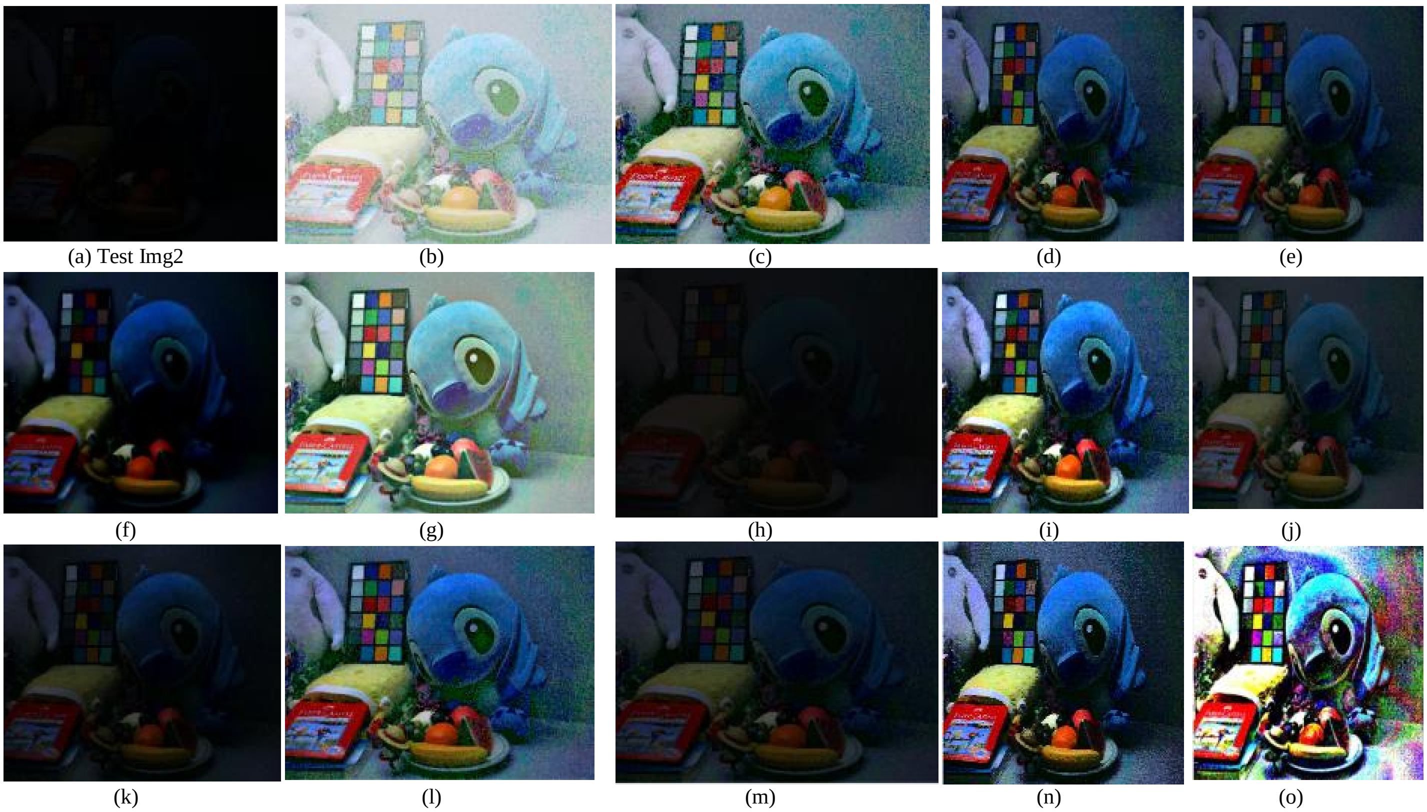

(a)Reference Image 1 (b)Gradient Magnitude of Reference image (c)Gradient Magnitude of Noisy Image

Fig. 7. Gradient Magnitude of Reference Image1 and Noisy Image.

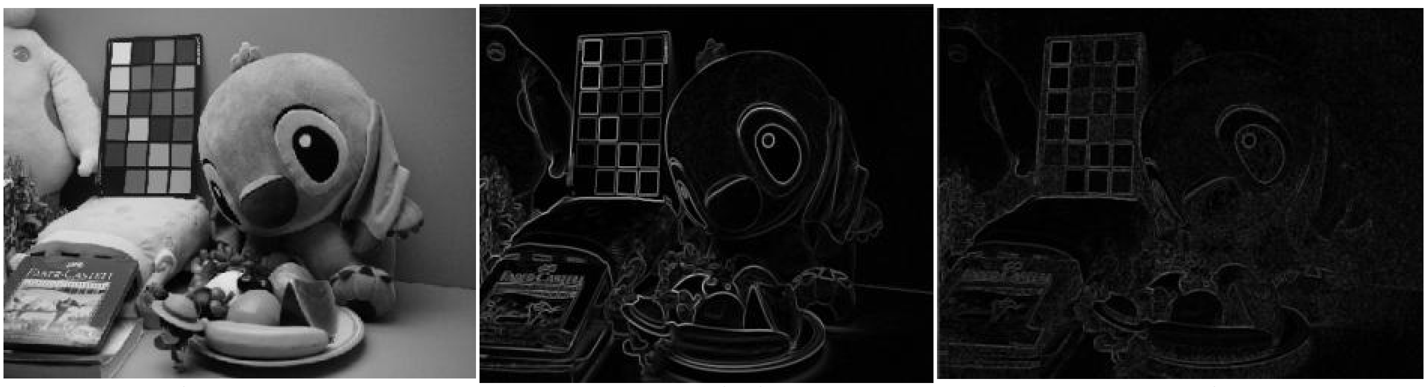


(a) Test Img2 (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

Fig. 8. The Comparison Outcomes Test Image2 (a) Real Low-light Image; The following Images are enhanced by: (b) IT2FB [20], (c) RBMA[26], (d) FBEM [27], (e) AIEM [28], (f) FEAE [29], (g) MSRA [30], (h) CAVIEHE [31], (i) LIME [32], (j) BIMEF [33], (k) LNET [34], (l) NPEA[35] [m] SRIE [36] [n]BPHE[37] [o] MSRCR [38].



(a)Reference Image 2 (b)Gradient Magnitude of Reference image (c)Gradient Magnitude of Noisy Image

Fig. 9. Gradient Magnitude of Reference Image2 and Noisy Image.

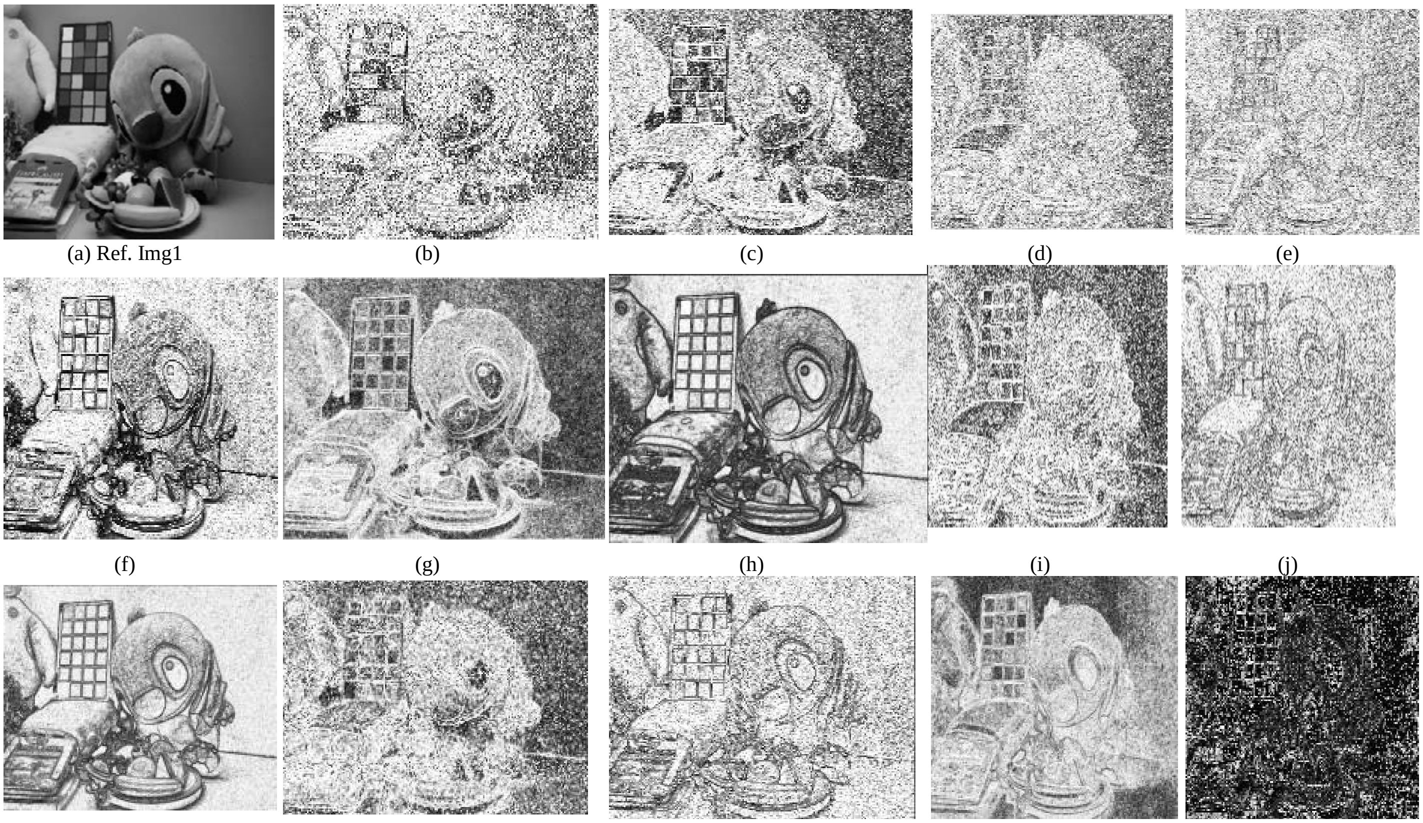


Fig. 10. The Comparison Outcomes Ref. Image1 (a) Reference Image; The following Images are enhanced by: (b) IT2FB [20], (c) RBMA[26], (d) FBEM [27], (e) AIEM [28], (f) FEAE [29], (g) MSRA [30], (h) CAVIEHE [31], (i) LIME [32], (j) BIMEF [33], (k) LNET [34], (l) NPEA[35] [m] SRIE [36] [n]BPHE[37] [o] MSRCR [38].

TABLE I. THE RECORDED MSE SCORES FOR THE COMPARATIVES (LOWEST SCORE IS THE BEST)

Image	**IT2FS**	**RBMP**	**FBEM**	**AIEM**	**FEAE**	**LIME**	**BIMEF**	**SRIE**
TestImg1	0.2957783	0.0933671	0.0285043	0.0764384	0.0578257	0.1863743	0.0378418	0.0137192
TestImg2	0.5330835	0.1811075	0.0440756	0.0213309	0.0116832	0.1065525	0.0159348	0.0069661
TestImg3	0.4311686	0.1884765	0.0257584	0.0761735	0.0856139	0.1251811	0.0475861	0.0212036
TestImg4	0.3250936	0.1169117	0.0256007	0.1254773	0.0850932	0.2084432	0.0461444	0.0220777

Image	**NPEA**	**BPHE**	**CAVIEHE**	**MSRA**	**MSRCR**	**LNET**
TestImg1	0.1084480	0.0078458	0.0429760	0.2861314	0.2885862	**0.0045625**
TestImg2	0.0749098	0.0845492	**0.0009261**	0.3079275	0.2841193	0.0029567
TestImg3	0.0694492	0.0267625	0.0333089	0.3367357	0.3580889	**0.0080669**
TestImg4	0.1519172	0.0239108	0.0644361	0.3583796	0.3300625	**0.0045718**

TABLE II. THE RECORDED RMSE SCORES FOR THE COMPARATIVES (LOWEST SCORE IS THE BEST)

Image	**IT2FS**	**RBMP**	**FBEM**	**AIEM**	**FEAE**	**LIME**	**BIMEF**	**SRIE**
TestImg1	0.5438550	0.3055603	0.1688320	0.2764750	0.2404697	0.4317110	0.1945297	0.1171289
TestImg2	0.7301257	0.4255673	0.2099419	0.1460511	0.1080888	0.3264238	0.1262332	0.0834629
TestImg3	0.6566343	0.4341388	0.1604943	0.2759956	0.2925986	0.3538094	0.2181423	0.1456145
TestImg4	0.5701698	0.3419235	0.1600022	0.3542278	0.2917075	0.4565558	0.2148126	0.1485858

Image	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.3293145	0.0885765	0.2073065	0.5349125	0.5372022	**0.0675460**
TestImg2	0.2736965	0.2907734	**0.0304320**	0.5549121	0.5330284	0.0543757
TestImg3	0.2635322	0.1635924	0.1825072	0.5802893	0.5984053	**0.0898158**
TestImg4	0.3897656	0.1546313	0.2538426	0.5986481	0.5745107	**0.0676148**

TABLE III. THE RECORDED PSNR SCORES FOR THE COMPARATIVES (HIGHEST SCORE IS THE BEST)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	53.4551363	58.4628604	63.6157018	59.3316821	60.5435907	55.4609380	62.3850827	66.7915145	57.8125822	69.2184311	61.8325401	53.5991442	53.5620441	**71.5728084**
TestImg2	50.8968467	55.5854348	61.7228179	64.8747043	67.4891830	57.8891633	66.1413308	69.7349266	59.4194148	58.8937050	**78.4981783**	53.2803149	53.6297926	73.4567085
TestImg3	51.8183279	55.4122275	64.0556064	59.3467575	58.8393542	57.1894109	61.3900027	64.9007048	59.7481264	63.8895391	62.9391988	52.8919083	52.6248905	**69.0977402**
TestImg4	53.0447151	57.4862196	64.0822808	57.1791464	58.8658485	54.9749219	61.5236057	64.7252557	56.3487290	64.3788502	60.0735098	52.6213668	52.9788370	**71.5639629**

TABLE IV. THE RECORDED WPSNR SCORES FOR THE COMPARATIVES (HIGHEST SCORE IS THE BEST)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	11.3176642	16.3443338	21.5312272	17.2358092	18.3756584	13.3539544	20.2668132	24.6841020	15.7201659	27.1490633	19.7112954	11.4881722	14.0373919	**29.4550061**
TestImg2	8.7625898	13.5024273	19.6510197	22.7892871	25.2990697	15.7966793	24.0579747	27.6668179	17.3674225	16.8182650	**36.3993706**	11.1827298	14.5294831	31.3554380
TestImg3	9.6761380	13.2787880	31.7233984	17.2394699	16.7057388	15.0955838	19.2632333	22.7914569	17.6404362	21.9613413	20.8354621	10.7732827	12.5262427	**26.9895351**
TestImg4	10.9035732	15.3530057	21.9893547	15.0685158	16.7149919	12.8668475	19.3981699	22.6181768	14.2447339	22.3875860	17.9577614	10.5024942	13.5243776	**29.5050471**

TABLE V. THE RECORDED SSIM SCORES FOR THE COMPARATIVES (HIGHEST SCORE IS THE BEST)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.99785	0.99523	0.98434	0.99278	0.98891	0.99682	0.98747	0.97937	0.97678	0.97538	0.98724	**0.99913**	0.98008	0.97369
TestImg2	0.98168	**0.99975**	0.99264	0.98822	0.98320	0.99736	0.98685	0.98275	0.99597	0.99509	0.97725	0.99625	0.98523	0.97966
TestImg3	0.97072	0.99230	0.99961	0.99918	0.99866	0.99692	**0.99994**	0.99934	0.99932	0.99870	0.99982	0.98041	0.98312	0.99775
TestImg4	0.99031	**0.99981**	0.99570	0.99967	0.99922	0.99714	0.99822	0.99520	0.99893	0.99338	0.99891	0.98940	0.98824	0.98988

TABLE VI. The Recorded CW-SSIM Scores for the Comparatives (Highest Score is the Best)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.74861	0.93698	0.86088	0.94114	0.93022	0.89871	0.88115	0.74399	0.82446	0.76114	0.94778	**0.95962**	0.40466	0.67500
TestImg2	0.69474	**0.96894**	0.81053	0.68657	0.72691	0.95084	0.57326	0.41762	0.78879	0.93743	0.18714	0.95356	0.42400	0.39128
TestImg3	0.85948	0.94236	0.96641	0.90545	0.86561	0.86520	**0.96931**	0.94301	0.88723	0.77239	0.97376	0.80165	0.25599	0.90355
TestImg4	0.83388	**0.97886**	0.93521	0.91537	0.85188	0.77317	0.96155	0.86016	0.86229	0.94870	0.97293	0.84677	0.41263	0.79696

TABLE VII. The Recorded VIF Scores for the Comparatives (Highest Score is the Best)

Image	IT2FB	RBMB	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.19582	0.46423	0.48409	0.79706	0.81646	2.48645	0.38585	0.30754	1.03547	0.29112	0.65896	**1.82329**	0.15908	0.17925
TestImg2	0.19125	0.71589	0.34668	0.16834	0.12851	0.85126	0.09787	0.06581	0.46828	0.71736	0.00814	**1.09521**	0.09469	0.03815
TestImg3	0.30926	1.05753	0.90069	1.39086	1.72082	3.19192	0.72948	0.73963	1.35298	1.82509	0.98186	**4.11126**	0.63433	0.44912
TestImg4	0.22262	0.56422	0.63636	1.41722	1.47031	3.51728	0.61541	0.59447	1.82540	1.14454	1.32809	**2.89510**	0.33520	0.31034

TABLE VIII. The Recorded UQI Scores for the Comparatives (Highest Score is the Best

Image	IT2FB	RBMB	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.50275	0.56244	0.36204	0.48627	0.46478	0.47524	0.44116	0.27996	0.48282	0.19056	0.41242	**0.55411**	0.00088	0.16764
TestImg2	0.24627	**0.37394**	0.28067	0.19893	0.11726	0.34546	0.17189	0.10729	0.33060	0.27312	0.02443	0.35001	-0.00078	0.06231
TestImg3	0.39627	0.55050	0.66431	0.60912	0.66688	0.50212	**0.69900**	0.68206	0.60753	0.40463	0.65395	0.36701	0.00070	0.56662
TestImg4	0.46861	**0.61940**	0.51680	0.52719	0.63598	0.43620	0.60771	0.51228	0.48266	0.33248	0.58218	0.41976	0.00084	0.31957

TABLE IX. The Recorded IEF Scores for the Comparatives

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE
TestImg1	14.02	6.80	2.18	4.61	3.08	9.00	2.70	1.66
TestImg2	1.38	**36.40**	3.50	2.21	1.56	8.85	1.98	1.52
TestImg3	0.21	0.81	15.15	7.19	4.62	1.92	**66.31**	9.30
TestImg4	1.51	**55.29**	3.58	28.64	17.87	4.36	8.47	3.22

Image	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	6.06	1.39	2.68	**23.19**	0.86	1.30
TestImg2	6.04	5.07	1.15	6.04	0.72	1.29
TestImg3	8.60	4.37	29.63	0.31	0.17	2.84
TestImg4	10.63	2.27	13.12	1.32	0.43	1.54

TABLE X. THE RECORDED IMMSE SCORES FOR THE COMPARATIVES (LOWEST SCORE IS THE BEST)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE
TestImg1	0.29578	0.09337	0.02850	0.07644	0.05783	0.18637	0.03784	0.01372
TestImg2	0.53308	0.18111	0.04408	0.02133	0.01168	0.10655	0.01593	0.00697
TestImg3	0.43117	0.18848	0.02576	0.07617	0.08561	0.12518	0.04759	0.02120
TestImg4	0.32509	0.11691	0.02560	0.12548	0.08509	0.20844	0.04614	0.02208

Image	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.10845	0.00785	0.04298	0.28613	0.28859	**0.00456**
TestImg2	0.07491	0.08455	**0.00093**	0.30793	0.28412	0.00296
TestImg3	0.06945	0.02676	0.03331	0.33674	0.35809	**0.00807**
TestImg4	0.15192	0.02391	0.06444	0.35838	0.33006	**0.00457**

TABLE XI. THE RECORDED MSSIM SCORES FOR THE COMPARATIVES (HIGHEST SCORE IS THE BEST)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE
TestImg1	0.69711	0.57042	0.61017	0.47714	0.47798	0.29046	0.64615	0.71292
TestImg2	0.49734	0.33369	0.48442	0.60775	0.70339	0.34148	0.67126	0.75477
TestImg3	0.86260	0.75360	0.82702	0.72642	0.66499	0.59373	0.84162	0.84437
TestImg4	0.80867	0.70498	0.69829	0.50175	0.48395	0.34765	0.74615	0.72605

Image	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.39561	0.79282	0.56256	0.33490	0.04524	**0.86121**
TestImg2	0.39457	0.39503	**0.95229**	0.29305	0.04378	0.84538
TestImg3	0.70558	0.61072	0.78073	0.55215	0.05001	**0.90839**
TestImg4	0.40896	0.58566	0.58085	0.38086	0.03150	**0.88989**

TABLE XII. THE RECORDED MAE SCORES FOR THE COMPARATIVES (LOWEST SCORE IS THE BEST)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE
TestImg1	.54073	0.28875	0.14273	0.24516	0.19210	0.35733	0.17784	0.09783
TestImg2	0.72464	0.38529	0.17624	0.12211	0.06624	0.26447	0.11178	0.07188
TestImg3	0.65597	0.42867	0.15295	0.26613	0.27547	0.33508	0.21354	0.13846
TestImg4	0.56909	0.33592	0.14793	0.33768	0.26626	0.41688	0.20788	0.13908

Image	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.28849	0.06415	0.17191	0.49991	0.32936	**0.05246**
TestImg2	0.23937	0.21595	**0.02866**	0.52285	0.30042	0.04276
TestImg3	0.25134	0.13097	0.17295	0.56538	0.43635	**0.08162**
TestImg4	0.37019	0.10905	0.23189	0.57767	0.38217	**0.05915**

TABLE XIII. The Recorded IQI Scores for the Comparatives (Highest Score is the Best)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.50275	**0.56244**	0.36204	0.48627	0.46478	0.47524	0.44116	0.27996	0.48282	0.19056	0.41242	0.55411	0.00088	0.16764
TestImg2	0.24627	**0.37394**	0.28067	0.19893	0.11726	0.34546	0.17189	0.10729	0.33060	0.27312	0.02443	0.35001	-0.00078	0.06231
TestImg3	0.39627	0.55050	0.66431	0.60912	0.66688	0.50212	**0.69900**	0.68206	0.60753	0.40463	0.65395	0.36701	0.00070	0.56662
TestImg4	0.46861	0.61940	0.51680	0.52719	**0.63598**	0.43620	0.60771	0.51228	0.48266	0.33248	0.58218	0.41976	0.00084	0.31957

TABLE XIV. The Recorded FSIM Scores for the Comparatives (Highest Score is the Best)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.99089	**0.99812**	0.99317	0.99212	0.99144	0.99439	0.99806	0.99367	0.98372	0.98823	0.99312	0.99382	0.89235	0.99166
TestImg2	0.98012	0.99134	0.98658	0.98997	0.97504	0.98937	0.99204	0.98644	0.98348	0.97946	**0.99228**	0.98625	0.85214	0.98798
TestImg3	0.98496	0.98844	0.99123	0.98093	0.98042	0.99171	**0.99419**	0.99001	0.98111	0.97666	0.99254	0.98894	0.87291	0.98457
TestImg4	0.99346	0.99454	0.98943	0.98195	0.98285	0.99169	**0.99713**	0.98671	0.97952	0.97625	0.99633	0.99078	0.88741	0.99364

TABLE XV. The Recorded EME Scores for the Comparatives (Lowest Score is the Best)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	**2.59748**	8.50280	15.49097	13.75025	12.11895	15.35547	10.66295	15.38244	14.20973	17.29908	11.87166	10.27982	6.84198	15.63085
TestImg2	**2.50348**	12.83307	14.85539	14.46187	11.29116	15.08368	13.58085	14.86042	14.29100	16.61080	8.13913	9.38662	7.12783	15.13008
TestImg3	**1.1726**	3.41829	7.41943	6.23491	5.58065	7.42239	4.84588	7.09565	6.87650	15.08932	7.21423	5.83274	11.2897	7.50607
TestImg4	**1.57324**	4.42487	9.87638	8.28385	7.23544	9.67597	6.20972	9.56697	9.05361	19.59307	8.44812	7.32125	9.20486	9.88048

TABLE XVI. The Recorded BRISQUE Scores for the Comparatives (Lowest Score gives Best Result)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE
TestImg1	**19.16849**	26.79686	28.17226	40.12349	36.34488	31.19316	24.76338	27.94361
TestImg2	38.52422	41.81477	39.58667	38.65845	46.63950	41.56760	37.20722	32.93015
TestImg3	**9.30332**	23.98699	29.88640	31.15363	26.12095	34.16256	22.76009	22.46018
TestImg4	**13.33258**	27.60852	27.55335	39.86039	24.63390	32.46380	24.94812	25.84723

Image	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	32.13035	27.45103	29.33331	28.04724	43.45818	29.61410
TestImg2	41.17648	40.32965	**20.65364**	41.47551	43.45818	36.82453
TestImg3	32.11621	35.37021	29.18756	34.10248	43.45818	20.30180
TestImg4	34.58516	26.89832	29.00054	26.91344	43.45818	27.04023

TABLE XVII. The Recorded NIQE Scores for the Comparatives (Lowest Score gives Best Result)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE
TestImg1	6.87128	7.69587	7.78807	9.30102	**4.77781**	8.65645	7.26530	6.83867
TestImg2	8.85253	9.66738	9.74843	9.22545	**4.51997**	9.62537	8.55000	7.83046
TestImg3	5.48614	6.63134	6.61409	7.71980	**3.35362**	6.89375	6.35760	5.99619
TestImg4	6.28973	7.31890	8.17693	10.12868	**5.31512**	9.25940	7.41547	7.41176

Image	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	8.31067	7.00298	7.80821	8.38040	38.50599	6.34044
TestImg2	9.95520	9.33642	6.45676	9.98429	30.40184	7.27924
TestImg3	6.83652	6.90076	6.91004	7.41921	29.21334	5.64895
TestImg4	9.26341	7.93860	8.48241	8.85040	35.71108	6.64925

TABLE XVIII. The Recorded PIQE Scores for the Comparatives (Low Score gives the Best Result

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE
TestImg1	26.42034	32.16037	**7.78807**	41.84593	32.86864	42.84311	26.99954	20.84821
TestImg2	50.00359	60.81876	55.35629	47.18139	44.25742	58.32593	40.86070	33.82644
TestImg3	**16.83733**	25.82195	29.24633	35.53923	26.52064	44.85693	21.50740	20.89448
TestImg4	**16.57803**	26.45110	31.64320	43.47681	18.68615	47.58669	24.56774	28.06270

Image	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	42.58162	16.01919	27.02401	42.62543	89.59252	11.76635
TestImg2	59.50193	55.91969	**31.78153**	62.50870	91.13846	20.94066
TestImg3	39.86982	48.48490	32.36529	46.31817	90.00533	17.35385
TestImg4	46.15295	38.91951	36.72036	44.24773	93.80261	11.13506

TABLE XIX. The Recorded SCC Scores for the Comparatives (Highest Score gives Best Result)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE
TestImg1	0.94437	0.96385	0.94473	0.94454	0.94763	0.93273	0.96627	0.93963
TestImg2	0.89702	0.93781	0.92440	0.92948	0.85067	0.91688	0.94332	0.91973
TestImg3	0.90259	0.92460	0.95753	0.90317	0.89467	0.96179	0.96355	0.94373
TestImg4	0.93555	0.94975	0.93647	0.90874	0.92014	0.94416	0.96413	0.92855

Image	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.87421	0.90175	**0.95252**	0.90587	0.03840	0.92522
TestImg2	0.89164	0.89831	**0.95326**	0.87015	0.02420	0.91458
TestImg3	0.87771	0.87153	**0.95395**	0.90131	0.01632	0.93460
TestImg4	0.82583	0.89521	**0.97046**	0.89288	-0.02249	0.94405

TABLE XX. The Recorded CVSI Scores for the Comparatives

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.06229	0.01437	0.03058	0.03095	0.04164	0.03950	0.01836	0.04131	0.04705	0.10903	0.05969	0.03166	**0.13403**	0.09452
TestImg2	0.12598	0.03197	0.04038	0.04257	0.13080	0.05623	0.02703	0.03347	0.05480	0.09028	0.03139	0.06614	**0.15729**	0.07505
TestImg3	0.11777	0.04921	0.02574	0.04486	0.04792	0.02469	0.02188	0.02985	0.04386	0.05375	0.03633	0.04424	**0.14413**	0.04457
TestImg4	0.08689	0.03801	0.05750	0.07298	0.07639	0.07271	0.02013	0.03138	0.09107	0.09174	0.03113	0.08400	**0.15153**	0.06126

TABLE XXI. The Recorded MCSD Scores for the Comparatives Score:Degree of Distortion-least gives Best Result

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.00004	0.00002	0.00002	0.00003	**0.00001**	0.00008	0.00002	0.00003	0.00002	0.00004	0.00002	0.00003	0.00051	0.00005
TestImg2	0.00006	**0.00001**	0.00004	0.00006	0.00009	0.00003	0.00008	0.00002	0.00003	0.00003	0.00014	0.00002	0.00054	0.00011
TestImg3	0.00002	**0.00001**	0.00002	0.00002	0.00002	0.00003	0.00004	0.00003	0.00002	0.00002	0.00002	0.00007	0.00057	**0.00001**
TestImg4	0.00003	0.00001	0.00002	0.00003	0.00002	0.00007	0.00001	0.00001	0.00002	0.00002	0.00001	0.00006	0.00057	0.00002

TABLE XXII. The Recorded NQM Scores for the Comparatives (Least Score Gives the Best) (Reference and Denoise)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	5.72822	10.90554	9.18275	13.25518	12.31878	5.95895	8.26158	6.81954	10.05271	5.74568	10.79705	10.65045	-4.19388	**4.89679**
TestImg2	5.21723	14.94767	7.43430	5.31561	5.02503	11.64804	4.12727	3.28481	8.25344	9.38074	1.68181	11.24193	-2.71410	**3.14435**
TestImg3	4.50081	8.47573	13.19743	6.02085	4.20663	1.85097	14.00744	10.91840	**3.55751**	1.97678	11.71896	1.03840	-12.47818	9.22515
TestImg4	4.78090	12.59009	9.60334	7.75522	4.82721	0.69335	12.12942	9.03722	**4.68356**	5.96796	11.60276	1.77181	-8.87400	6.73111

TABLE XXIII. The Recorded GMSM Scores for the Comparatives (Higher the Score gives Good Quality

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE
TestImg1	0.7755	0.8126	0.8048	0.7657	0.8248	0.6780	**0.8284**	0.8050
TestImg2	0.7179	0.6582	0.7187	0.7539	0.6794	0.6614	**0.7597**	0.7455
TestImg3	0.8649	0.8753	0.8897	0.8404	0.8837	0.7385	**0.9106**	0.9069
TestImg4	0.8275	0.8406	0.8262	0.7037	0.8458	0.5951	**0.8599**	0.8377

Image	**NPEA**	**BPHE**	**CAVIEHE**	**MSRA**	**MSRCR**	**LNET**
TestImg1	0.7182	0.7511	0.8094	0.7114	0.3007	0.7517
TestImg2	0.6626	0.6718	0.6316	0.6228	0.2666	0.7165
TestImg3	0.8213	0.6868	0.8698	0.7111	0.3000	0.8970
TestImg4	0.6381	0.7080	0.7877	0.6315	0.3038	0.8507

TABLE XXIV. THE RECORDED GMSD SCORES FOR THE COMPARATIVES

Image	**IT2FB**	**RBMP**	**FBEM**	**AIEM**	**FEAE**	**LIME**	**BIMEF**	**SRIE**
TestImg1	0.8806	0.9015	0.8971	0.8751	0.9082	0.8234	**0.9101**	0.8972
TestImg2	0.8473	0.8113	0.8477	0.8683	0.8242	0.8132	**0.8716**	0.8634
TestImg3	0.9300	0.9356	0.9432	0.9167	0.9400	0.8594	**0.9542**	0.9523
TestImg4	0.9097	0.9169	0.9089	0.8389	0.9197	0.7714	**0.9273**	0.9153

Image	**NPEA**	**BPHE**	**CAVIEHE**	**MSRA**	**MSRCR**	**LNET**
TestImg1	0.8475	0.8666	0.8997	0.8434	0.5483	0.8670
TestImg2	0.8140	0.8197	0.7947	0.7892	0.5163	0.8465
TestImg3	0.9062	0.8287	0.9327	0.8433	0.5478	0.9471
TestImg4	0.7988	0.8414	0.8875	0.7946	0.5511	0.9223

TABLE XXV. THE RECORDED AG SCORES FOR THE COMPARATIVES (HIGHER THE SCORE GIVES GOOD QUALITY)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE
TestImg1	0.10996	0.15797	0.15792	0.21128	0.14827	0.33065	0.13524	0.11716
TestImg2	0.13219	0.22952	0.14650	0.09817	0.03967	0.20956	0.07968	0.06527
TestImg3	0.06604	0.11923	0.11339	0.14094	0.11490	0.20601	0.09822	0.09938
TestImg4	0.07574	0.11887	0.13126	0.20760	0.15153	0.30025	0.11576	0.12324

Image	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.24745	0.10121	0.02154	0.30972	0.78260	0.08131
TestImg2	0.19189	0.18526	0.02162	0.27452	1.08063	0.04200
TestImg3	0.14384	0.18717	0.11866	0.22959	0.85070	0.07664
TestImg4	0.24331	0.17288	0.17302	0.28082	0.94841	0.07655

TABLE XXVI. THE RECORDED CONTRAST SCORES FOR THE COMPARATIVES (HIGHER THE SCORE GIVES GOOD QUALITY)

Image	**IT2FB**	**RBMP**	**FBEM**	**AIEM**	**FEAE**	**LIME**	**BIMEF**	**SRIE**
TestImg1	**0.07470**	0.04234	0.02362	0.03674	0.02989	0.05101	0.02811	0.01785
TestImg2	**0.09490**	0.05139	0.02453	0.01756	0.01028	0.03579	0.01622	0.01108
TestImg3	**0.09416**	0.06499	0.02966	0.04412	0.04521	0.05306	0.03742	0.02774
TestImg4	**0.07958**	0.04960	0.02541	0.04981	0.04052	0.05981	0.03320	0.02438

Image	**NPEA**	**BPHE**	**CAVIEHE**	**MSRA**	**MSRCR**	**LNET**
TestImg1	0.04230	0.01353	0.01243	0.06943	0.04037	0.01205
TestImg2	0.03262	0.02960	0.01678	0.06900	0.03820	0.00734
TestImg3	0.04213	0.0268	0.0248	0.08262	0.05564	0.02049
TestImg4	0.05392	0.02023	0.02123	0.08055	0.04698	0.01408

TABLE XXVII. THE RECORDED IE SCORES FOR THE COMPARATIVES (HIGHER THE SCORE GIVES GOOD QUALITY)

Image	IT2FB	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	15	**15.87423**	15.66459	15.75404	15.54836	15.63918	15.80531	15.64746	15.73456	15.49749	15.37739	15.83352	-14.71573	-15.56715
TestImg2	**15.98439**	15.80274	15.53808	15.53808	15.30713	15.57657	15.74564	15.68848	15.72043	15.52732	15.42732	15.82062	-14.49740	-15.57494
TestImg3	**15.99621**	15.97231	15.89987	15.93132	15.88706	15.89043	15.94712	15.89795	15.91559	15.64596	15.54596	15.93316	-14.87595	-15.86741
TestImg4	**15.99558**	15.97075	15.79935	15.92798	15.88523	15.89030	15.94342	15.89896	15.92207	15.57102	15.43202	15.94793	-14.50907	-15.85952

TABLE XXVIII. RUN TIME FOR ALL ALGORITHMS

Images	IT2FS	RBMP	FBEM	AIEM	FEAE	LIME	BIMEF	SRIE	NPEA	BPHE	CAVIEHE	MSRA	MSRCR	LNET
TestImg1	0.345445	0.379292	6.800890	7.571706	3.968177	0.163928	**0.256611**	17.412172	9.307077	0.262238	0.912000	1.343199	2.893714	8.809306
TestImg2	0.362200	0.3528	14.746129	6.500533	1.919185	0.148863	**0.233173**	12.652796	9.362168	0.194414	0.478139	1.264595	2.800885	5.862136
TestImg3	0.199488	**0.196766**	23.093706	6.120637	1.797364	0.180001	0.223248	6.430840	9.325342	0.221893	0.732831	1.166960	2.739701	6.661681
TestImg4	0.203637	**0.164310**	13.216258	7.291177	1.810248	0.143545	0.250914	9.901286	10.921751	0.289308	1.464919	1.209983	2.707332	7.045211

TABLE XXIX. RUN TIME EVALUATION OF ALL ALGORITHMS

Sl.No	Algorithm	Run Time (sec) 400x600	Run Time (sec) 701x1052	Run Time (sec) 3264x2175
1	IT2FB	0.350821	0.641391	6.538643
2.	RBMA	0.318945	0.7460	4.922583
3.	FBEM	12.994208	18.771755	92.741121
4.	AIEM	4.134789	6.267442	10.84552
5.	FEAE	1.953480	3.367680	12.701537
6.	LIME	**0.146787**	**0.529567**	**4.343078**
7.	BIMEF	0.215428	0.582480	5.241805
8.	SRIE	17.259832	25.738757	695.323839
9.	NPEA	9.267901	28.419136	427.278078
10.	BPHE	0.304741	0.598509	5.122004
11.	CAVIEHE	0.974524	2.187483	35.554106
12.	MSRA	1.354881	2.706997	17.385311
13	MSRCR	2.815200	8.631183	150.519389
14	LNET	7.112936	15.468000	45.789045

VI. Conclusion and Future Work

Low Light image enhancement formulas are more helpful for various vision applications. It can be found that many of the existing scientific study have neglected a lot of issues; i.e. no technique is precise for different circumstances. The review has demonstrated the undeniable fact that shown methods have neglected the methods to reduce the noise concern which can be shown within the output images of the image enhancement algorithms. The issue of uneven and also over illumination may also be an issue for enhancement methods. So it will be expected to change the prevailing methods in this manner that altered strategy may continue steadily to function better. In near future, to eliminate the issues of present research a different integrated algorithm is going to be proposed.

Table XXX shows the performance evaluation of all Algorithms. In this paper fourteen Image enhancement algorithms were compared and finally LNET gives the best output quality image and LIME method gives the least run time.

TABLE XXX. Performance Evaluation of All Algorithms

Sl.No	**QM**	**TestImg1**	**TestImg2**	**TestImg3**	**TestImg4**
1	MSE	LNET	CAVIEHE	LNET	LNET
2.	RMSE	LNET	CAVIEHE	LNET	LNET
3.	PSNR	LNET	CAVIEHE	LNET	LNET
4.	WPSNR	LNET	CAVIEHE	LNET	LNET
5.	SSIM	MSRA	RBMP	BIMEF	RBMP
6.	CW-SSIM	MSRA	RBMP	BIMEF	RBMP
7.	VIF	MSRA	MSRA	MSRA	MSRA
8.	UQI	MSRA	BIMEF	MSRA	RBMB
9.	IEF	MSRA	RBMP	BIMEF	RBMP
10.	IMMSE	LNET	CAVIEHE	LNET	LNET
11.	MSSIM	LNET	CAVIEHE	LNET	LNET
12.	MAE	LNET	CAVIEHE	LNET	LNET
13	IQI	RBMP	RBMP	BIMEF	FEAE
14	FSIM	RBMP	CAVIEHE	BIMEF	BIMEF
15	EME	IT2FB	IT2FB	IT2FB	IT2FB
16	BRISQUE	IT2FB	CAVIEHE	IT2FB	IT2FB
17	NIQE	FEAE	FEAE	FEAE	FEAE
18	PIQE	FBEM	CAVIEHE	IT2FB	IT2FB
19	SCC	CAVIEHE	CAVIEHE	CAVIEHE	CAVIEHE
20	CVSI	MSRCR	MSRCR	MSRCR	MSRCR
21	MCSD	FEAE	RBMP	RBMP	LNET
22	NQM	LNET	LNET	NPEA	NPEA
23	GMSM	BIMEF	BIMEF	BIMEF	BIMEF
24	GMSD	BIMEF	BIMEF	BIMEF	BIMEF
25	AG	MSRCR	MSRCR	MSRCR	MSRCR
26	C	IT2FB	IT2FB	IT2FB	IT2FB
27	IE	RBMP	IT2FB	IT2FB	IT2FB

Acknowledgment

Our heartfelt thanks goes to the Management and the Principal of CITech for providing infrastructure and a wholehearted support to conduct research at CITech.

References

[1] Lee, C. Lee, C.S. Kim, Contrast enhancement based on layered difference representation of 2d histograms, IEEE Trans. Image Process. 22 (12) (2013) 5372–5384.

[2] Guo, X., Li, Y., Ling, H. (2016). LIME: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26(2): 982-993. https://doi.org/10.1109/TIP.2016.2639450.

[3] Wang, Y.F., Liu, H.M., Fu, Z.W. (2019). Low-light image enhancement via the absorption light scattering model. IEEE Transactions on Image Processing, 28(11): 5679-5690. https://doi.org/10.1109/TIP.2019.2922106.

[4] Park, S., Yu, S., Kim, M., Park, K., Paik, J. (2018). Dual autoencoder network for retinex-based low-light image enhancement. IEEE Access, 6: 22084-22093. https://doi.org/10.1109/ACCESS.2018.2812809.

[5] Dai, C., Lv, Y., Long, Y., Sui, H. (2018). A novel image enhancement technique for tunnel leakage image detection. Traitement du Signal, 35(3-4): 209-222. https://doi.org/10.3166/TS.35.209-222.

[6] Jung, C., Yang, Q., Sun, T., Fu, Q., Song, H. (2017). Low light image enhancement with dual-tree complex wavelet transform. Journal of Visual Communication and Image Representation, 42: 28-36. https://doi.org/10.1016/j.jvcir.2016.11.001.

[7] Kim, W., Lee, R., Park, M., Lee, S.H. (2019). Low-light image enhancement based on maximal diffusion values. IEEE Access, 7: 129150-129163. https://doi.org/10.1109/ACCESS.2019.2940452.

[8] Thung, K.-H. and Raveendran, P. (2009) A Survey of Image Quality Measures. IEEE Technical Postgraduates (TECHPOS) International Conference , Kuala Lumpur, 14-15 December 2009, 1-4.

[9] Jobson D, Rahman Z (1997) Properties and performance of a center/surround retinex. IEEE Trans Image Process A Publ IEEE Signal Process Soc 6(3):451–462.

[10] Jobson DJ, Rahman Z, Woodell GA (2002) A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Trans Image Process 6(7):965–976.

[11] Rahman Z, Jobson DJ, Woodell GA (2004) Retinex processing for automatic image enhancement.J Electron Imaging 13(1):100–110.

[12] Zhengang S, Liqun G, Kun W (2007) A novel approach to image enhancement and thresholding based on fuzzy theory. In: IEEE Conference on industrial electronics and applications 2201–2205.

[13] Kong XW (2007) The fuzzy image enhancement algorithm for iow snr image. Laser J 5:44–45.

[14] Z.Ying, G. Li, and W. Gao, “A Bio-Inspired Multi-Exposure Fusion Framework for Low-light Image Enhancement,” arXiv:1711.00591 [cs], Nov. 2017.

[15] Z.Ying, G. Li, Y. Ren, R. Wang, and W. Wang, “A New Image Contrast Enhancement Algorithm Using Exposure Fusion Framework,” in International Conference on Computer Analysis of Images and Patterns, 2017, pp. 36–46.

[16] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, “Deep bilateral learning for real-time image enhancement,” ACM Trans. Graph., vol. 36, no. 4, pp. 1–12, Jul. 2017.

[17] K. G. Lore, Adedotun Akintayo, and S. Sarkar, “LLNet: A deep autoencoder approach to natural low-light image enhancement,” Pattern Recognition, vol. 61, pp. 650–662, Jan. 2017.

[18] Abdullah-Al-Wadud M , Kabir M H , Dewan M A A , et al. A Dynamic Histogram Equalization for Image Contrast Enhancement[J]. IEEE Transactions on Consumer Electronics, 2007, 53(2):p.593-600.

[19] Chulwoo Lee, Chul Lee, and Chang-Su Kim, “Contrast enhancement based on layered difference representation of 2D histograms,” IEEE Transactions on Image Processing, vol. 22, no. 12, pp. 5372-5384, Dec. 2013.

[20] Chaira, T. (2014). An improved medical image enhancement scheme using Type II fuzzy set. Applied Soft Computing, 25: 293-308. https://doi.org/10.1016/j.asoc.2014.09.004.

[21] Zohair Al-Ameen (2021)“Contrast Enhancement of Digital Images Using an Improved Type-II Fuzzy Set-Based Algorithm”,International Information and Engineeirng Technology Association,38:39-50 https://doi.org/10.18280/ts.380104.

[22] Zou, Y., Dai, X., Li, W., Sun, Y. (2015). Robust design optimisation for inductive power transfer systems from topology collection based on an evolutionary multi-objective algorithm. IET Power Electronics, 8(9): 1767-1776. https://doi.org/10.1049/iet-pel.2014.0468.

[23] Kallel, F., Sahnoun, M., Hamida, A.B., Chtourou, K. (2018). CT scan contrast enhancement using singular value decomposition and adaptive gamma correction. Signal, Image and Video Processing, 12(5): 905-913. https://doi.org/10.1007/s11760-017-1232-2.

[24] Jobson, D.J., Rahman, Z.U., Woodell, G.A. (1997). Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing, 6(3): 451-462. https://doi.org/10.1109/83.557356.

[25] Hanumantharaju, M.C., Ravishankar, M., Rameshbabu, D.R. (2013). Design and FPGA implementation of an 2D Gaussian surround function with reduced on-chip memory utilization. In 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Mysore, pp. 604-609. https://doi.org/10.1109/ICACCI.2013.6637241.

[26] Mohammad Abid Al-Hashim, Zohair Al-Ameen(2020) Retinex-Based Multiphase Algorithm for Low-Light Image Enhancement International Information and Engineeirng Technology Association 37:733-743, https://doi.org/10.18280/ts.370505.

[27] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding and John Paisley, A Fusion-based Enhancing Method for Weakly Illuminated Images, Signal Processing, http://dx.doi.org/10.1016/j.sigpro.2016.05.031.

[28] Wencheng Wang, Zhenxue Chen, Xiaohui Yuan, Xiaojin Wu, “Adaptive Image Enhancement Method for Correcting Low-Illumination Images”, Information Sciences-2019, https://doi.org/10.1016/j.ins.2019.05.015.

[29] Dong, X., G. Wang, Y. Pang, W. Li, J. Wen, W. Meng, and Y. Lu. "Fast efficient algorithm for enhancement of low lighting video." Proceedings of IEEE® International Conference on Multimedia and Expo (ICME). 2011, pp. 1–6.

[30] Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. 2017, 26, 982–993. [CrossRef] [PubMed].

[31] Ying, Z.; Li, G.; Gao, W. A bio-inspired multi-exposure fusion framework for low-light image enhancement. arXiv, 2017; arXiv:1711.00591.

[32] Fu, X.; Zeng, D.; Huang, Y.; Zhang, X.P.; Ding, X. A weighted variational model for simultaneous reflectance and illumination estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2782–2790.

[33] Wang, S.; Zheng, J.; Hu, H.-M.; Li, B. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Trans. Image Process. 2013, 22, 3538–3548. [CrossRef] [PubMed].

[34] Contrast Enhancement Using Brightness Preserving Bi- Histogram Equalization, YEONG-TAEKGI M, 1997 IEEE.

[35] S. Palanikumar1,*, M. Sasikumar2, J. Rajeesh3 Entropy Optimized Palmprint Enhancement Using Genetic Algorithm and Histogram Equalization, International Journal of Genetic Engineering 2012, 2(2): 12-18 DOI: 10.5923/j.ijge.20120202.01.

[36] Daniel J. Jobson, Member, IEEE, Zia-ur Rahman, Member, IEEE, and Glenn A. WoodellA Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 6, NO. 7, JULY 1997.

[37] Jinxiang Maa,b, Xinnan Fana,c,d, Jianjun Nic,d , Xifang Zhub,* and Chao XiongbMSRCR Image Enhancement Based on Gaussian Filtering and Guided Filtering, International Journal of Modern Physics B • May 2017, DOI: 10.1142/S0217979217440775.

[38] Chongyi Li, Jichang Guo, Fatih Porikli, Yanwei Pang, LightenNet: a Convolutional Neural Network for weakly illuminated image enhancement, Pattern Recognition Letters (2018), doi: 10.1016/j.patrec.2018.01.010.

Effect of Feature Engineering Technique for Determining Vegetation Density

Yuslena Sari[1]
Doctoral Department of Agricultural Science
Universitas Lambung Mangkurat
Banjarmasin, Indonesia

Yudi Firmanul Arifin[2]
Faculty of Forestry
Universitas Lambung Mangkurat
Banjarmasin, Indonesia

Novitasari Novitasari[3]*
Faculty of Engineering
Universitas Lambung Mangkurat
Banjarmasin, Indonesia

Mohammad Reza Faisal[4]
Department of Computer Science
Universitas Lambung Mangkurat
Banjarmasin, Indonesia

***Abstract*—Vegetation density is one type of information collected from vegetation cover. Vegetation density influences evapotranspiration in terrain, which is essential in assessing how vulnerable peatlands are to fire. The Keetch and Byram Drought Index model, which evaluates peatland fire vulnerability, divides vegetation density into heavily grazed, softly grazed, and un-grazed. Manual approaches for analyzing vegetation density in the field, on the other hand, need a significant amount of resources. Image data acquisition, pre-processing, feature extraction, classification, feature selection, classification, and validation are all computer vision approaches used to solve these problems. Artificial intelligence algorithms and machine learning approaches promise outstanding accuracy in modern computer vision research. However, in the classification process, the impact of feature extraction is critical. Pattern identification at Back Propagation Neural Network (BPNN) is problematic because the feature extraction dimension is excessively complicated. The solution to this problem is using the feature engineering technique to choose the characteristics. This research aims to explore how feature engineering influences the accuracy of results. According to the statistics, implementing the recommended strategy can increase accuracy by 1% and increase kappa by 1.5%. This increase in vegetation density classification accuracy might help detect peatland vulnerability sooner. The novel aspect of this paper is that, after feature extraction, a feature engineering strategy is used in the machine learning classification stage to reduce the number of complex dimensions.**

***Keywords*—Vegetation cover; vegetation density; feature extraction; feature engineering; accuracy**

I. Introduction

The Keetch and Byram Drought Index (KBDI) model uses vegetation cover conditions as one of the parameters to generate the land drought index [1]–[5]. The density of vegetation cover in KBDI peat is classified into three groups: extensively grazed, gently grazed, and un-grazed. To quantify peatland fire vulnerability used the KBDI peat drought index[3]. Peatland fires have occurred in South Kalimantan in recent years. The fires have become a more common occurrence. Whether intended or not, human activities are responsible for 99.9% of land fires. The existence of human intervention is referred to as anthropogenic [6]. Automation can predict land fire susceptibility to anthropogenic influences. Land cover analysis can be automated utilising machine-learning artificial intelligence approaches combined with computer vision techniques.

Various researches have attempted to classify vegetation cover using computer vision and artificial intelligence techniques. The precision and accuracy of vegetation cover identification have greatly improved because of advancements in remote sensing data (Dronova et al., 2012; Mihail et al., 2018; Rios et al., 2021). Land classification analysis using remote sensing, on the other hand, necessitates a significant investment of time and money on the part of researchers.

Previous research indicated that using ordinary cameras for automated land cover categorisation could not discern between trees, weeds, and grass. They have not accomplished these difficulties [7]–[9]. Artificial intelligence-based machine learning classification is necessary to distinguish between trees, weeds, and grasses. The classification method consists of a support vector machine (SVM), naive Bayes, and an artificial neural network (ANN) [10]–[13]. According to some of that research, image classification requires feature extraction or the conversion of an image into numerical to be classified using artificial intelligence methods. Features are crucial in the field of image classification and recognition.

On the other hand, previous researchers looked at converting natural colour information (RGB) to more suited colour space and used that to discern between vegetation and non-vegetation. Philipp and Rath [14] distinguish between vegetation and non-vegetation using the Lab, Luv, and HSV colour spaces. However, because the dimensions of feature extraction are too broad and complicated, machine learning techniques perform poorly. By lowering the dimensions of feature extraction, feature engineering can minimise the number of input features [15], [16].

To deal with the challenge of distinguishing between heavily grazed, lightly grazed, and un-grazed. A novel

*Corresponding Author.

technique was developed that combined pre-processing, feature extraction and selection, as well as classification. Segmentation based on distance threshold is employed as pre-processing. We employed grey level concurrent matrix (GLCM) and Backward Elimination to extract and choose the feature. A backpropagation neural network is utilised in the classification approach or analytical methods to distinguish between heavily grazed, lightly grazed, and ungrazed. Based on the approach, a technique for acquiring wetland image data will be developed and tested on pure vegetation stands with a height > 20 m above the canopy [12].

The contribution of this study was validated by comparing the results of feature selection based on the classification method to (1) show the effect of feature selection on classification performance; (2) analyze the relationship between feature selection and image classification performance; (3) determine the parameters needed to achieve the best performance results on the classification model, and (4) present the results of the best classification model for determining vegetation density.

This paper's contents are categorized as follows: Previous works on this research are highlighted in Section II. Section III outlines our research strategy. Section IV explains the materials and methods used in our experiment. The experiment design is described in Section V. Following that Section V summarizes the experiment's results and discussion. Finally, in Section VI, we reach a conclusion.

II. Related Work Use

There have been previous land categorisation studies in general land regions [17]–[19], wetlands [20]–[24], and peatland research [25], [26]. The study by Herdiyeni et al. [8] looked at how to identify plant health by utilising characteristics based on the local binary pattern variant (LBPV) approach, morphological features, and colour features and employed a probabilistic neural network (PNN) with a performance of 72.15%.

The bulk of this research employed artificial intelligence to evaluate land cover using remote sensing data, while field validation was rarely performed. For land cover analysis, the most often used artificial intelligence approaches include supervised machine learning, Back-propagation neural network (BPNN) with an accuracy of 97.65%, support vector machine (SVM) with an accuracy of 97.45%, and neural networks (ANN) with an accuracy of 96.95% [17]. The study of Tan et al. [18] compares the performance of the random forest (RF), decision tree (DT), SVM, and ANN techniques to map three typical landscapes. The results show that ANN performs relatively poorly compared to the performance of other methods. Based on this research, further research is needed to improve the ANN method, which is done in this research. In addition, based on Zaldo-Aubanell et al. [19] stated that this research is important and highly preferred, especially when applying data nationally, and has been carried out by many world researchers to solve their domestic problems. This is the reason for the importance of using data that follows the problems faced nationally.

III. Material and Proposed Method

A. Tools and Location Research

The study took place in Block I of the Liang Anggang protected forest in Banjarbaru, South Kalimantan. Kayu Tangi Production Forest Management Unit [27] is in charge of this area. Fig. 2 depicts the research site. Fig. 2(a) research site in data collecting and Fig. 2(b) location distance between Syamsuddin Noor airstrips illustrate the research location. The nearby population has primarily utilised the existing peatland for agricultural land and plants. Peatlands ranging from shallow (100 — 200 cm) to extremely deep (> 300 cm) encompass 749.87 hectares (78.43 %). According to Zaldo-Aubanell et al. [19] this data based on a national scale and more important also widely used because it adapts more uses to the regional national scale.

B. Proposed Framework

Land cover picture data from the research location was utilised to create the dataset for this investigation. The data used was 450 images taken from drones 20 meters above the item. At this point, the proposed technique processes variables extracted from feature extraction using the Backward Elimination algorithm. Feature selection picks features that influence the BPNN classification process accuracy. The hidden layer, neuron size, momentum, learning rate, and training cycle are BPNN indicators. Measurements are used to determine whether or not the categorization findings are accurate and optimum (Fig. 1).

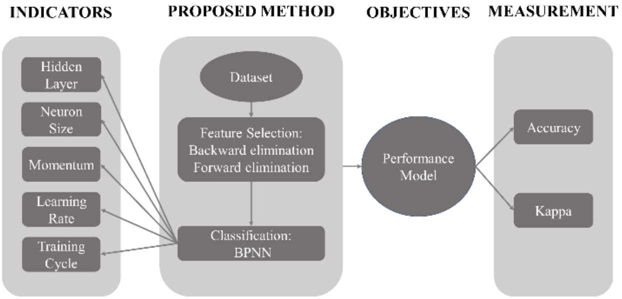


Fig. 1. Research Framework.

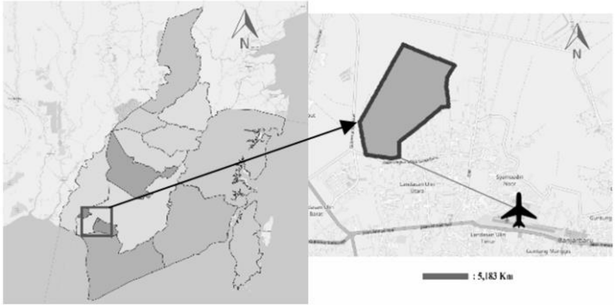


Fig. 2. The Research Location (a) The Location on the Map of South Kalimantan in the Red Box and (b) The Distance from the Syamsuddin Noor Airstrip.

C. Research Design

This research used artificial intelligence based on BPNN as a computer vision methodology to categorise vegetation density. The procedure consists of data capture, pre-processing, segmentation, feature extraction, feature selection, classification, and validation are the steps of classification design. Fig. 3 illustrates the steps.

1) Data acquisition: With a 900-degree gimbal angle, data was collected using a Mavic Pro drone. The drone is positioned at a distance of 20 meters from the object. The information is separated into three categories: heavily grazed, moderately grazed, and ungraded. The three groups will be separated based on the plant growth at the location of the study. Herbs and shrubs that are lightly grazed make up tree vegetation in heavily grazed areas. On the other hand, un-grazed consists of dry (dead) vegetation, rivers (water), land, and towns. Up to 300 images are collected for use as training data.

2) Pre-processing image: This research focuses on vegetation density, and initial image processing is done using a static threshold value. Then segmentation is done using the green colour of the tree as a threshold. This study used the Euclidean distance (equation 1) approach to colour distance-based segmentation.

$$d(x,y) = \sqrt{(x-y)^2} \tag{1}$$

3) Feature extraction: The chosen image is converted into a set of numerical parameters. This numerical parameter is critical for differentiating an item. Using a Grey-Level Co-occurrence Matrix (GLCM), the feature extraction approach is employed to get quantitative values. The likelihood of two grey levels co-occurring is stored in GLCM [28].

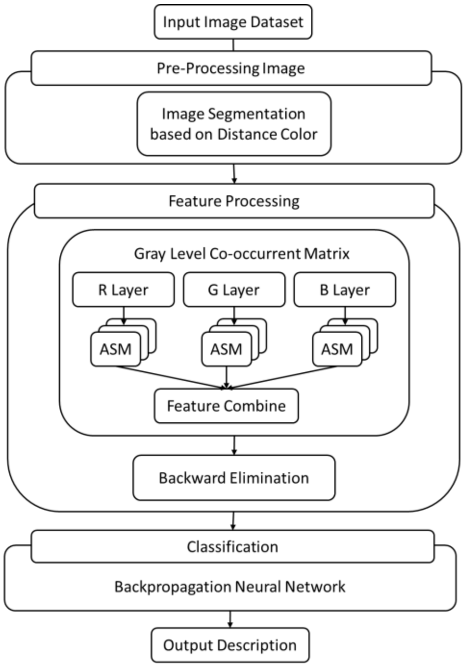


Fig. 3. Research Design.

The second-order moment or energy (ene), entropy (ent), contrast (con), homogeneity (hom), and correlation are retrieved from the vegetation picture to represent the data co-occurrence matrices indicated in equations (2) to (6) (cor). The distribution of co-occurrence values is designated by k and l with different angles, 0^0, 45^0, 90^0, and 135^0, at the offset provided (1,1) by p(k,l). Rotation invariant, mean, and variance of orientation-dependent characteristics are computed individually for different angles using different angles.

$$ene = \sum_{k=0}^{L-1}\sum_{l=0}^{L-1}\{P(k,l)\}^2 \tag{2}$$

$$ent = -\sum_{k=0}^{G-1}\sum_{l=0}^{G-1}P(k,l)\times Log\big(P(k,l)\big) \tag{3}$$

$$con = \sum_{n=0}^{G-1}n^2\{\sum_{k=0}^{G}\sum_{l=0}^{G}P(k,l)\}\{n = |k-l| \tag{4}$$

$$hom = \sum_{k=0}^{G-1}\sum_{l=0}^{G-1}\frac{1}{1+(k-l)^2}P(k,l) \tag{5}$$

$$cor = \frac{\sum_{k=0}^{G-1}\sum_{l=0}^{G-1}(k,l)(P(k,l) - \mu_k{'}\mu_l{'}}{\sigma_k{'}\sigma_l{'}} \tag{6}$$

The equation is used to solve the correlation equation (7).

$$\begin{aligned}
P_x(k) &= \sum_{l=0}^{G-1}P(k,l)\\
P_y(l) &= \sum_{k=0}^{G-1}P(k,l)\\
\mu'_k &= \sum_{k=0}^{G-1}\sum_{l=0}^{G-1}k * P(k,l)\\
\mu'_l &= \sum_{k=0}^{G-1}\sum_{l=0}^{G-1}l * P(k,l)\\
\sigma'_k &= \sum_{k=0}^{G-1}\sum_{l=0}^{G-1}P(k,l)(k-\mu'_l)^2\\
\sigma'_l &= \sum_{k=0}^{G-1}\sum_{l=0}^{G-1}P(k,l)(l-\mu'_l)^2
\end{aligned} \tag{1}$$

4) Feature selection: Feature Selection is a machine learning method in which a collection of data features is utilised to train algorithms. Feature selection has become a hot topic in pattern recognition, statistics, and data mining, according to Oded Maimon [29].

One of the essential aspects that might affect classification accuracy is feature selection since if the dataset contains multiple features, the dataset's dimensions will be significant, lowering classification accuracy. The issue with feature selection is dimensionality reduction, as all elements are required initially to achieve maximum accuracy.

According to Maimon [29], there are four fundamental causes for dimension reduction:

- Decreasing the learning cost.
- Increasing the learning performance.
- Reducing outside dimensions.
- Reducing redundant dimensions.

Because not all features/attributes are relevant to the problem, the fundamental notion of Feature Selection is to choose a subset of existing characteristics without transforming them. Some of these qualities or attributes are even bothersome and diminish accuracy. To increase accuracy, noisy or useless features must be deleted. Furthermore, having many characteristics or qualities will slow down the computation process. Backward elimination starts with the entire collection of characteristics and removes any leftover features from the specified ExampleSet in each round. Performance is calculated for each element published using the inner operator, such as cross-validation. Only the attributes that cause a modest performance decrease are finally eliminated from consideration. Then a new round with a different selection begins. This method removes the usage of additional memory, the memory used to hold the data, and any memory necessary to perform the inner operator. After the termination requirements are fulfilled, the speculative spin parameter defines how many spins will be made in a row. Elimination will continue if performance improves during the theoretical round. Any extra missing characteristics would be restored if no speculative spin were performed. This process might help avoid the model being stuck on a local optimum.

The difference with forwarding selection is that it starts with an empty attribute and adds any new characteristics from the specified set in each round. The inner operator, such as cross-validation, is used to assess performance for each feature added. Only the most significant performance boost attribute is included in the selection. Then a new round with a different selection begins.

5) Classification: According to Witten [30], data mining is a series of processes to obtain knowledge or patterns from data sets. Data mining solves the problem by analysing the data already in the database. The method of finding a model or function that explains or distinguishes a concept or data class intends to estimate the course of an object whose label is unknown [31].

The classification stage uses the BPNN artificial intelligence method. An artificial neural network (ANN) is a learning algorithm that implements a simple network connected to neurons and units. The performance of ANN is similar to the version of the human brain in recognising a specific pattern. ANN can provide effective classification results even though the input data contains noise and is incomplete [32], [33]. One type of ANN that is often used is Backpropagation. Backpropagation architecture comprises three layers, including the input layer, hidden layer and output layer [32], [33]. Fig. 4 shows a simple visualisation of the Backpropagation structure. Backpropagation formulation can be formulated in equation 8.

$$y = f(\sum_{i=1}^{L} w_i x + b) \qquad (2)$$

By displaying the weight vector, x as the input vector, and b as the bias value, f is the activation function. One hidden layer (10 neurons), one output layer, purely log sig activation function, and 1000 epochs were utilised.

6) Validation: Validation measures are utilised to assess the classification performance of the model. This study uses two types of validation measures: accuracy and kappa. The percentage of cells categorised exactly in class I to the total number of cells is called accuracy. The following is an example of accuracy:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} x\ 100\% \qquad (9)$$

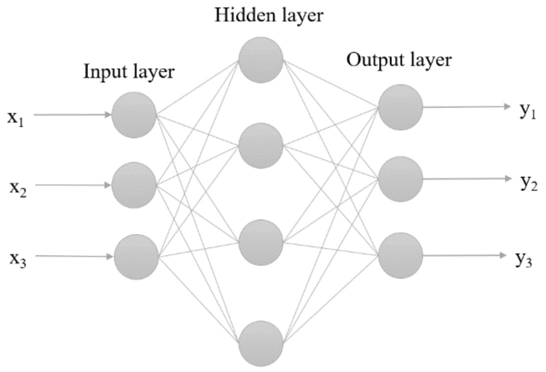


Fig. 4. Backpropagation Architecture.

True Negative (TN), False Positive (FP), False Negative (FN), and True Positive (TP) values will be calculated to obtain accuracy, precision and recall values. The accuracy value describes how accurately the system can classify the data correctly. In other words, the accuracy value compares the data that is classified correctly and the whole data. The precision value describes the number of positive data categories classified correctly divided by the total data classified as positive.

Another indicator of accuracy is the kappa coefficient. Kappa is a measure of how the result of a classification compares to a given value by chance. It can take a deal from 0 to 1. If the kappa coefficient equals 0, there is no similarity between the classified image and the reference image. If the kappa coefficient equals 1, the image is classified, and the ground truth image is identical. Thus, the higher the kappa coefficient, the more accurate is the classification.

IV. Experimental Design

This study demonstrates how data influences the best characteristics. The approach used on this dataset is tested during the classification process, ensuring that the classification results are accurate and based on the best features. The accuracy gained in this study is compared to other experiments to determine the optimum accuracy.

The original image is used to extract features using the feature extraction approach, eq. (2) to (6) using the GLCM feature extraction method. For each image, a total of 60 features are created. These attributes are taken from each colour layer, which includes red, green, and blue. According to equations (1) to (3), GLCM creates 20 texture-based features for each colour received from four distinct angles, with each corner yielding five particular features.

Consequently, the final data consists of 300 photos multiplied by 60 characteristics. The BPNN classification algorithm was used to analyse the feature findings. In this study, each parameter in the BPNN classification technique was evaluated, and an assessment was based on the usage of the backward elimination-based feature selection approach. Backward elimination findings are compared to past investigations, specifically forward selection.

In performance evaluation, sampling and validation processes are employed with multiple tens. The dataset is divided into ten values for cross-validation, with each component evenly distributed. The experiment was repeated ten times according to the number of cross-validations and the average training and performance classification testing outcomes. The classification model's performance is assessed when a confusion matrix is created. This investigation obtained the most notable performance findings using MATLAB (www.mathworks.com).

V. Result and Discussion

The discussion is broken down into three stages: data preparation, feature extraction, and classification (Fig. 2). At each stage of the research, the results are revealed.

A. Segmentation

The segmentation results show segmentation that only takes the green colour from the image. To obtain the image's unique characteristics according to the agreed class will then use the segmented image. Table I is an image of the segmentation results of the three types of labels that have been agreed upon.

B. Feature Extraction

The GLCM technique is used to turn the selected pictures into functional characteristics. A new matrix with 300 images and 60 features is created by combining feature approaches. From Table II each row of the matrix represents data, while the columns reflect the characteristics of each piece of information.

The column comprises the variables F1, F2, F3, F4 to F120, as given in Table I. Columns are variables that hold the values for each GLCM texture feature. The GLCM method's second-moment angle features, contrast, and correlation with 0 degrees orientation direction on the red layer are F1, F2, F3, and F4. In the GLCM approach, the other Fn characteristics include orientation and colour layers. The kurtosis value on the blue layer is the next F60 feature. The data numbers received from the dataset are represented by the matrices in rows 1 to 300. All characteristics were employed in the classification process in this study. Therefore it's safe to infer that they all help get the best classification results.

C. Evaluation of BPNN Classification based on Parameters

The supporting parameters are used to evaluate the BPNN classification. The test results are shown in Fig. 5, utilising a training cycle value of 10 to 100 cycles.

As shown in the Fig 5, the results indicate that the optimal training cycle was used for 100 cycles with an accuracy of 84.67% and a kappa of 77%. These findings suggest that increasing the number of training cycles increases accuracy and kappa performance. This result is consistent with numerous other studies that employed neural network training cycle settings and showed excellent performance on long training cycles.

The learning rate experiment on the BPNN technique is also applied from 0.1 to 1, utilising the highest performance results from the training cycle, which is 100 cycles (Fig. 6). The results reveal that a learning rate of 0.1 produces the most significant outcomes compared to other learning rates. When applying a high learning rate, the pattern created from the results of the accuracy performance based on the learning rate reveals that performance steadily falls. The solution is not ideal because the high learning rate and performance are reduced [34].

TABLE I. Segmentation Result

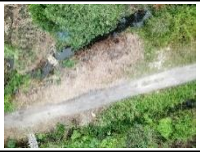
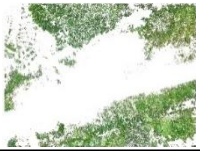
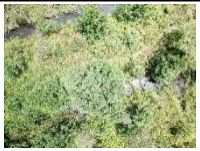
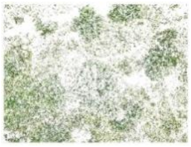
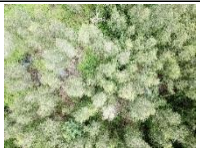
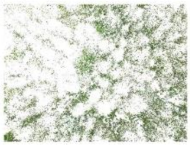
Label	Data image	Segmented image
un-grazed		
softly grazed		
heavily grazed,		

TABLE II. Feature Extraction Applied to Original Data

No	F1	F2	F3	F4	…	F60
1	1.E-04	631.25	0.08	9.73	…	3.E-04
2	1.E-04	449.81	0.10	9.67	…	3.E-04
3	2.E-04	309.74	0.11	9.22	…	6.E-04
4	2.E-04	302.13	0.11	9.20	…	6.E-04
…	…	…	…	…	…	…
…	…	…	…	…	…	…
299	9.E-05	576.55	0.09	9.76	…	3.E-04
300	8.E-05	630.58	0.09	9.85	…	3.E-04

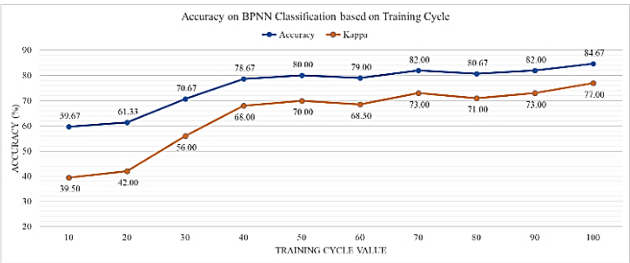


Fig. 5. Evaluation of BPNN based on Training Cycle.

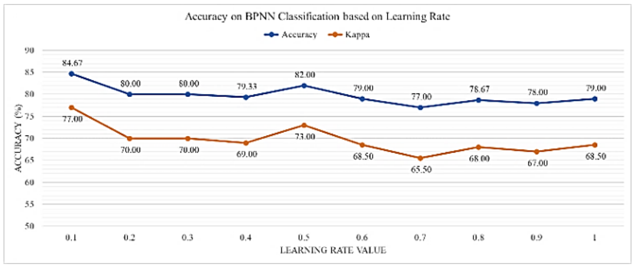


Fig. 6. Evaluation of BPNN based on Learning Rate.

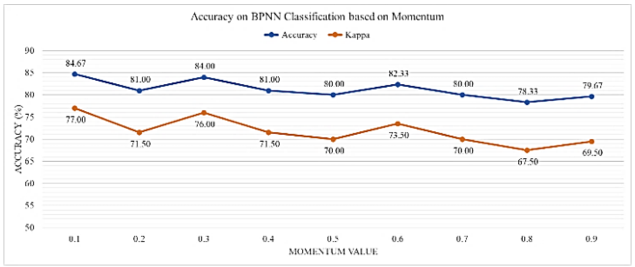


Fig. 7. Evaluation of BPNN based on Momentum.

The momentum parameter is used to evaluate BPNN in Fig. 7. The accuracy performance based on momentum follows the same pattern as the learning rate results, with ups and downs, and the most significant results when given a momentum value of 0.1. These findings align with earlier research that claims that a small momentum value brings features closer together without boosting convergence [34].

D. BPNN Classification Testing based on Feature Selection

To choose features in this experiment used the backward elimination feature. The maximum number of eliminations parameters is somewhere between 10 and 60 (shown in Fig. 8). These parameters are applied to the BPNN classification algorithm, utilising the optimal training cycle, learning rate, and momentum parameters from the previous experiment, 100, 0.1, and 0.1, respectively.

The accuracy is 85.67 % when using a maximum total elimination of 10, and backward elimination training delivers significant results. These data show a 1% improvement over the BPNN classification algorithm without feature selection. With a 1.5 % increase, measurements with kappa have the same performance pattern as measurements with accuracy.

E. Evaluation Comparison between Feature Selection

We used the forward selection to acquire the best performance and compare assessments for feature selection (shown in Fig. 9). Forward selection is based on the selection of empty attributes, and each iteration adds unneeded attributes from the quantity of data for each additional attribute, according to the working idea. Only the traits that increase performance the most are added to the selection in the forward selection, which begins with the alteration of the selection. Backward elimination allows for the most significant results since it starts with a complete data set and deletes each characteristic for each repetition. Backward elimination, like the forward selection, uses cross-validation to predict performance and deletes attributes that cause a drop in performance.

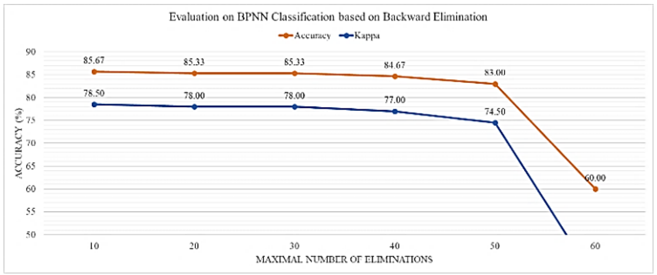


Fig. 8. Evaluation of BPNN based on Backward Elimination.

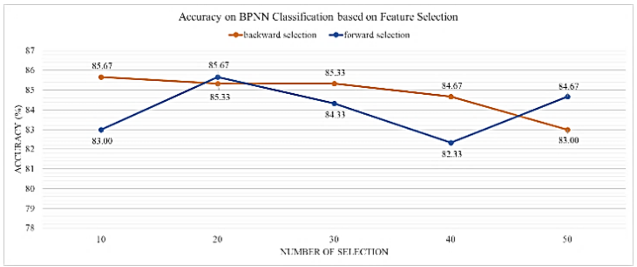


Fig. 9. Comparison BPNN Classification based on Features Selection.

The findings demonstrate that when ten features are applied, and the selected features are eliminated via backward selection, the most outstanding performance obtained from both feature selection approaches offers the same accuracy of 85.67%. Using forward selection produced an 85.67% performance with 20 and 60 chosen features, respectively. The backward selection highlights the behaviour of the accuracy findings as much as possible by picking features from a large number of characteristics in the data. It is conceivable that using the feature impacts the findings' correctness. While the forward selection is based on the calculation of the correlation matrix by taking the relationship between features that produce the highest correlation coefficient and only considering the relationship between features, the backward selection is based on the calculation of the correlation matrix by taking the relationship between features that produce the lowest.

VI. Conclusion

This study aims to determine the state of plant cover to quantify peatland fire vulnerability. As a result, this research may help some fire-prone countries overcome their problems. This research provides an automated identification approach based on the original image and ambient circumstances. The improved performance is due to the revised flow for determining cover situations. When the feature selection approach is coupled, the findings demonstrate an increase in performance—the proposed strategy results in a 1% improvement in accuracy and a 1.5 % increase in kappa. The rise happened when the feature selection utilised forward and backward elimination features. As a result, many features have suboptimal capabilities, and the feature selection approach can provide native features that are suitable for determining vegetation cover situations. Feature engineering can minimise the number of input dimensions in visual feature extraction while increasing the accuracy of vegetation density categorisation. As a result, it can improve machine learning. Using the Keetch and Byram Drought Index model, it will be more effective to use engineering characteristics in the vegetation density workflow classification system to evaluate peatland fire. The scope of this study is confined to making suggestions about the impact of feature engineering. At the classification step, to enhance accuracy may make further efforts by comparing machine learning classification approaches.

Acknowledgment

This study was completed with partial funding from Riset Keilmuan with the Riset Mandiri Dosen scheme from LPDP through the Directorate of Resources, the Directorate General of Higher Education, Research, and Technology. Ministry of Education, Culture, Research and Technology (contract: 010/E4.1/Ak.04.RA/2021). The authors thank to LPPM Universitas Lambung Mangkurat.

References

[1] P. Ganatsas, M. Antonis, and T. Marianthi, "Development of an adapted empirical drought index to the Mediterranean conditions for use in forestry," Agric. For. Meteorol., vol. 151, no. 2, pp. 241–250, 2011, doi: 10.1016/j.agrformet.2010.10.011.

[2] Garcia-Prats, D. C. Antonio, F. J. G. Tarcísio, and M. J. Antonio, "Development of a Keetch and Byram-Based drought index sensitive to forest management in Mediterranean conditions," Agric. For. Meteorol., vol. 205, pp. 40–50, 2015, doi: 10.1016/j.agrformet.2015.02.009.

[3] N. Novitasari, J. Sujono, S. Harto, A. Maas, and R. Jayadi, "Drought index for peatland wildfire management in central kalimantan, indonesia during el niño phenomenon," J. Disaster Res., vol. 14, no. 7, pp. 939–948, 2019, doi: 10.20965/jdr.2019.p0939.

[4] L. G. Liacos, "Soil Moisture Depletion in the Annual Grass Type," J. Range Manag., vol. 15, no. 2, p. 67, 1962, doi: 10.2307/3894863.

[5] H. a. J. Taufik, M., Setiawan, B. I. dan van Lanen, "Modification of a fire drought index for tropical wetland ecosystems by including water table depth," 2015.

[6] Hope, U. Chokkalingam, and S. Anwar, "The stratigraphy and fire history of the Kutai Peatlands, Kalimantan, Indonesia," Quat. Res., vol. 64, no. 3, pp. 407–417, 2005, doi: 10.1016/j.yqres.2005.08.009.

[7] S. Haug, A. Michaels, P. Biber, and J. Ostermann, "Plant classification system for crop /weed discrimination without segmentation," 2014 IEEE Winter Conf. Appl. Comput. Vision, WACV 2014, pp. 1142–1149, 2014, doi: 10.1109/WACV.2014.6835733.

[8] Y. Herdiyeni and M. M. Santoni, "Combination of morphological, local binary pattern variance and color moments features for Indonesian medicinal plants identification," 2012 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2012 - Proc., no. December, pp. 255–259, 2012.

[9] D. Kendal, C. E. Hauser, G. E. Garrard, S. Jellinek, K. M. Giljohann, and J. L. Moore, "Quantifying Plant Colour and Colour Difference as Perceived by Humans Using Digital Images," PLoS One, vol. 8, no. 8, pp. 1–11, 2013, doi: 10.1371/journal.pone.0072296.

[10] N. Dong, L. Zhao, J. Chang, and A. Wu, "Research on Feature Selection and Classification Recognition Algorithm of Cervical Cell Image," Hunan Daxue Xuebao/Journal Hunan Univ. Nat. Sci., vol. 46, no. 12, pp. 1–8, 2019, doi: 10.16339/j.cnki.hdxbzkb.2019.12.001.

[11] P. Petropoulos, C. Kalaitzidis, and K. Prasad Vadrevu, "Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery," Comput. Geosci., 2012, doi: 10.1016/j.cageo.2011.08.019.

[12] L. P. e. Silva, A. P. C. Xavier, R. M. da Silva, and C. A. G. Santos, "Modeling land cover change based on an artificial neural network for a semiarid river basin in northeastern Brazil," Glob. Ecol. Conserv., vol. 21, 2020, doi: 10.1016/j.gecco.2019.e00811.

[13] Sitthi, M. Nagai, M. Dailey, and S. Ninsawat, "Exploring land use and land cover of geotagged social-sensing images using naive bayes classifier," Sustain., 2016, doi: 10.3390/su8090921.

[14] Philipp and T. Rath, "Improving plant discrimination in image processing by use of different colour space transformations," Comput. Electron. Agric., 2002, doi: 10.1016/S0168-1699(02)00050-9.

[15] W. Zhang and F. Gao, "Performance analysis and improvement of naïve Bayes in text classification application," 2013 IEEE Conf. Anthol. Anthol. 2013, pp. 1–4, 2013, doi: 10.1109/ANTHOLOGY.2013.6784818.

[16] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," Remote Sens. Environ., vol. 221, no. March 2018, pp. 430–443, 2019, doi: 10.1016/j.rse.2018.11.032.

[17] H. Chughtai, H. Abbasi, and I. R. Karas, "A review on change detection method and accuracy assessment for land use land cover," Remote Sens. Appl. Soc. Environ., vol. 22, no. March, p. 100482, 2021, doi: 10.1016/j.rsase.2021.100482.

[18] Tan, J. Zuo, X. Xie, M. Ding, Z. Xu, and F. Zhou, "MLAs land cover mapping performance across varying geomorphology with Landsat OLI-8 and minimum human intervention," Ecol. Inform., vol. 61, p. 101227, 2021, doi: https://doi.org/10.1016/j.ecoinf.2021.101227.

[19] Q. Zaldo-Aubanell, I. Serra, J. Sardanyés, L. Alsedà, and R. Maneja, "Reviewing the reliability of Land Use and Land Cover data in studies relating human health to the environment," Environ. Res., vol. 194, p. 110578, 2021, doi: https://doi.org/10.1016/j.envres.2020.110578.

[20] M. Abalo, D. Badabate, F. Fousseni, W. Kpérkouma, and A. Koffi, "Landscape-based analysis of wetlands patterns in the Ogou River basin in Togo (West Africa)," Environ. Challenges, vol. 2, p. 100013, 2021, doi: https://doi.org/10.1016/j.envc.2020.100013.

[21] Bunyangha, M. J. G. Majaliwa, A. W. Muthumbi, N. N. Gichuki, and A. Egeru, "Past and future land use/land cover changes from multi-temporal Landsat imagery in Mpologoma catchment, eastern Uganda," Egypt. J. Remote Sens. Sp. Sci., 2021, doi: https://doi.org/10.1016/j.ejrs.2021.02.003.

[22] H. Su, W. Yao, Z. Wu, P. Zheng, and Q. Du, "Kernel low-rank representation with elastic net for China coastal wetland land cover classification using GF-5 hyperspectral imagery," ISPRS J. Photogramm. Remote Sens., vol. 171, pp. 238–252, 2021, doi: https://doi.org/10.1016/j.isprsjprs.2020.11.018.

[23] R. Í. Magnússon et al., "Shrub decline and expansion of wetland vegetation revealed by very high resolution land cover change detection in the Siberian lowland tundra," Sci. Total Environ., vol. 782, p. 146877, 2021, doi: https://doi.org/10.1016/j.scitotenv.2021.146877.

[24] D. Mao, Y. Tian, Z. Wang, M. Jia, J. Du, and C. Song, "Wetland changes in the Amur River Basin: Differing trends and proximate causes on the Chinese and Russian sides," J. Environ. Manage., vol. 280, p. 111670, 2021, doi: https://doi.org/10.1016/j.jenvman.2020.111670.

[25] M. Karlson, M. Gålfalk, P. Crill, P. Bousquet, M. Saunois, and D. Bastviken, "Delineating northern peatlands using Sentinel-1 time series and terrain indices from local and regional digital elevation models," Remote Sens. Environ., vol. 231, p. 111252, 2019, doi: https://doi.org/10.1016/j.rse.2019.111252.

[26] Räsänen and T. Virtanen, "Data and resolution requirements in mapping vegetation in spatially heterogeneous landscapes," Remote Sens. Environ., vol. 230, p. 111207, 2019, doi: https://doi.org/10.1016/j.rse.2019.05.026.

[27] Muhammad Abdul Qirom, T. A. Windawati, Kissinger Kissinger, and A. Fithria, "Carbon stock potential on various land covers in heath forest in Liang Anggang, South Kalimantan," J. Galam, vol. 1, no. 2, pp. 61–78, 2021, doi: 10.20886/glm.2021.1.2.61-78.

[28] S. Santosa, R. A. Pramunendar, D. P. Prabowo, and Y. P. Santosa, "Wood Types Classification using Back-Propagation Neural Network based on Genetic Algorithm with Gray Level Co-occurrence Matrix for Features Extraction," IAENG Int. J. Comput. Sci., vol. 46, no. 2, 2019.

[29] O. Maimon and L. Rokach, Data Mining and Knowledge Discovery Handbook. 2010.

[30] H. Witten, E. Frank, and M. A. Hall, Data Mining, Third. 2016.

[31] Rutledge, "The Top Ten Algorithms in Data Mining," J. Qual. Technol., vol. 41, no. 4, 2009, doi: 10.1080/00224065.2009.11917798.

[32] Y. Sari, P. B. Prakoso, and A. R. Baskara, "Application of neural network method for road crack detection," Telkomnika (Telecommunication Comput. Electron. Control., vol. 18, no. 4, pp. 1962–1967, 2020, doi: 10.12928/TELKOMNIKA.V18I4.14825.

[33] Y. Sari, H. Suhud, A. R. Baskara, R. A. Pramunendar, and I. F. Radam, "Parking Lots Detection in Static Image Using Support Vector Machine Based on Genetic Algorithm," Int. J. Intell. Eng. Syst., vol. 14, no. 6, pp. 476–487, 2021, doi: 10.22266/ijies2021.1231.42.

[34] R. A. Pramunendar, S. Wibirama, P. I. Santosa, P. N. Andono, and M. A. Soeleman, "A robust image enhancement techniques for underwater fish classification in marine environment," Int. J. Intell. Eng. Syst., 2019, doi: 10.22266/ijies2019.1031.12.

Techno-pedagogical Solution to Support the Improvement of the Quality of Education in Technical and Vocational Training in Mauritania

Cheikhane SEYED[1], Jeanne Roux NGO BILONG[2], Mohamed Ahmed SIDI[3], Mohamedade Farouk NANNE[4]
Laboratory LIRT, Polytechnic Superior School, University Cheikh Anta DIOP (UCAD), Dakar, SENEGAL[1, 2]
General Director of Society DCS-SARL, Nouakchott, Mauritania[3]
Faculty of Sciences and Technologies, University of Nouakchott Al Aasriya (UNA)[4]
Department of Mathematics and Computer Science[4]

***Abstract*—E-learning is the most promising and fastest growing activity since the advent of the COVID-19 pandemic. Although the pandemic seems to have been eradicated in several countries of the world, it is worth mentioning that some positive cases of the covid-19 variant have been detected. This could accelerate the rate of infection again. Hence the interest generate in reinforcing the quality of distance learning platforms. Technical and vocational training (TVT) in Mauritania is based on science, technology, engineering, and mathematics (STEM) disciplines. Unfortunately, the expansion of the COVID-19 pandemic has negatively impacted the quality of education with a halt in teaching affecting 8000 students. Yet, the quality of education in these disciplines is a key factor in meeting the demands of emergence and economic growth. This paper advocates the mixed pedagogical model by proposing a techno-pedagogical solution to improve the quality of teaching and learning processes. The proposed solution combines the use of technologies such as Modular Object-Oriented Dynamic Learning Environment (Moodle) and Web Real Time Communication (WebRTC) to provide pedagogical services in a context with a limited Internet connection. In addition, we set up a signaling system to maintain direct communication between the pairs, Application Programming Interface (API) of Multipoint Control Unit (MCU) to ensure simultaneous collaboration in a peer-to-peer context, used implementations of security protocols such as Datagram Transport Layer Security (DTLS) and Secure Real-time Transport Protocol (SRTP) to secure data transport.**

Keywords—TVT; STEM; Mixed education; Moodle; WebRTC; signaling system; MCU; DTLS; SRTP

I. INTRODUCTION

Education is a key instrument for promoting growth and economic development. In view of this, the Mauritanian government, with the support of its partners, has launched a project under the leadership of DCS-SARL for the development of the technical and vocational training sector (TVT).

The training centers in Mauritania welcome around of 8,000 students each year in Science, Technology, Engineering and Mathematics (STEM) disciplines. These disciplines require the implementation of strategies to ensure the quality of the teaching system in order to meet the requirements of economic development requiring the performance of the educational process.

In addition, the pandemic of the COVID-19 impacted the living and working conditions and has in turn forced the States to take measures concerning social distancing in order to prevent its propagation in the society. To this effect, these professional technical training (FTP) centers have been victims of a blockage in their teaching-learning processes, drastically reducing the number of class sessions. This has had serious consequences on both the pedagogical continuity and the quality of teaching.

In spite of these shortcomings, these training centers must work to improve their teaching in accordance with quality assurance standards to enable students to be competitive and effective in the labor market, but also to increase the ratio of the population of age to attend these professional training courses.

The lack of a suitable pedagogical model for these training centers and the lack of a better techno-pedagogical infrastructure have favored the migration to ICT-based solutions in order to improve the quality of teaching, while ensuring pedagogical continuity online.

The need for a pedagogical model adapted to these training centers and the poor quality of the techno-pedagogical infrastructure led us to encourage these training centers to migrate to ICT-based solutions in order to improve the quality of teaching, while ensuring pedagogical continuity online in case face-to-face courses are not possible.

In the literature, authors [1] [2] have shown the usefulness of adopting the blended learning model for traditional training. The latter consists in combining traditional teaching methods with e learning in order to increase the efficiency of the teaching systems.

Other authors [3] have proposed mechanisms to improve the effectiveness of blended learning by adopting the model of teaching based on the integration of online and offline activities (O2O). This strategy aims to increase the utilization of the teaching system by providing access to learning content in an environment with limited internet connection.

In the works [4] [5] [6], the authors have advocated the so-called social-constructivist pedagogical model to improve the quality of teaching, according to which the teacher creates learning situations that invite learners to collaborate and cooperate via innovative techno-pedagogical services in order to develop concrete professional skills.

Teaching must rest on professional practice to meet the needs of the labor market. The authors [7] have proposed strategies to involve synchronous education development in the learning process.

This form of education, involving instantaneous, face-to-face interaction, allows schools to train students in correct concepts and awareness of work and mastery of professional skills through specialized synchronous courses.

These research discussions led us to propose a quality hybrid solution using innovative and flexible technologies to the profile of TVT centers. The proposed solution offers synchronous and asynchronous pedagogical services appropriate and adapted to the context of the TVT centers.

The rest of this paper is organized as follow: Section II describes technical and vocational training and discusses of Technological background in e-learning; Section III presents software architecture of system proposed; Section IV describes hybrid learning architecture; Section V is dedicated to discussion and result and finally in Section VI provides the conclusion and perspectives of ours works.

II. Related Work

In [8], the authors propose a survey. This survey reveals the fear, the anxiety of learners for the handling of distance learning platforms.

The authors' contribution [9] presents the collaborative platforms Google Classroom, Zoom and Microsoft Team and compares them. They show the negative impact of using distance learning platforms on the mental health of learners.

These limitations have led us to propose a technological solution appropriate and suitable to the context of technical and vocational training.

A. Technical and Vocational Training

The technical and vocational training (TVT) in Mauritania is provided by the National Institute of Promotion [10]. Its purpose is to provide individuals with the knowledge and skills necessary to practice a trade or profession in order to integrate into the labor market as a worker's helper, specialized worker, skilled worker, technician or senior technician.

The mission of the TVT articulated around the following axes:

1) Satisfaction the needs of the labor market in qualified personnel.

2) Improving the professional skills of workers.

3) The development of the individual's potentialities in the perspective of the accomplishment of his professional project.

4) The promotion of the entrepreneurial spirit, with a view to self-employment.

5) Educational and professional orientation, information and advice on skills.

The Technical and Vocational Training system under the MENFTR composed of: (1) 16 technical education and vocational training schools (EETFP); (2) Higher Center for Technical Education (CSET) and (3) Six private training establishments that train mainly in the tertiary and service sectors.

Concerning the training curriculum, the TVT offers initial training courses organized as follows:

- Certificate of Competence (CC); 6 to 9 months duration.
- Certificate of Professional Aptitude (CAP), lasting 2 years.
- Technician's Certificate (BT), lasting 2 years.
- Higher Technician Certificate (BTS), lasting 2 years.
- Technical Education Certificate (BET), lasting 4 years.
- Technical baccalaureate, lasting 3 years.

The number of TVT students for the 2020-2021 training year is 7,885, including:

- 238 at CSET.
- 7,125 students in EETFPs.
- 190 students at the Ighraa Institute.
- 332 Pupils in Private Establishments.

These students divided according to educational level as follows:

- BTS : 722.
- BT : 2726.
- CAP : 2895.
- Technical baccalaureate and BET: 1542.

B. Technological Background in e-learning

Today, ICT have become an integral part of the culture of society. Their places are increasingly important in all areas of life. This presence of ICT is all the more felt at the level of education systems.

Several studies such as [11] [12] show that the appropriate use of technological innovations in teaching can contribute to improving the quality of teaching and learning.

Nevertheless, technical and professional training resting essentially on STEM disciplines requires the adoption of learning practices that promote interactivity and collaboration to develop practical and professional skills in students that will enable them to compete on the market of work.

In addition, the low level of TVT students in the field of ICT encourages the adoption of flexible technological solutions to use in order to eliminate any obstacle to the use of ICT to the detriment of the deep exploitation of educational services.

Indeed, the authors [13] have shown the relevance of using Moodle technology to facilitate interaction and create a dynamic learning environment for mixed-mode courses. According to the authors [14], Moodle is a practical tool containing versatile and rich enough functionalities for teaching.

Moodle remains an asset for learners because it offers the possibility of working at their own pace by alternating course sessions and face-to-face and distance learning activities. In this context, we use this technology to provide, through our techno-pedagogical solution, an adaptation and organization of lessons with the aim of developing a certain autonomy in the learner: we are talking about asynchronous learning.

In addition, asynchronous teaching via Moodle illustrates some insufficient of teaching when one is looking for real-time exchange. For the authors [15], face-to-face interaction between teaching actors is a key factor in improving the quality of online teaching.

This synchronous teaching modality reflects traditional classroom teaching where students become actively involved in their learning by interacting with each other, but also with their teacher to produce their educational activities.

Several technologies have materialized the synchronous mode in the online teaching process by promoting collaboration between actors.

However, online collaborative solutions applying in an organizational network and including interactive spaces in STEM fields require a certain type of access to resources. The generalization of broadband access is one of the fundamental conditions for quality online education.

Unfortunately, the context of technical and vocational training centers often has a limited internet connection. This can pose a problem of efficiency of the distance education system. Research work [16] on technologies that can operate in a local IP environment and offer innovative collaborative services via the web, led us to take an interest in WebRTC technology.

According to the analysis study in the works [17], WebRTC technology is better compare to other real-time interaction tools for several reasons: (1) WebRTC allows interaction in peer-to-peer mode without the using an intermediate server. This remarkably reduces latency; (2) WebRTC integrated into the bandboxes of recent browsers without any intermediate configuration or installation.

In view of these discussions, we combine these two technologies, namely Moodle and WebRTC, to provide a quality e-learning solution offering educational services in synchronous and asynchronous mode, with the possibility of operating in an intranet network. In addition, these services are accessible via a simple browser requiring neither download nor installation of plugins on client workstations. This could have a positive impact on the implementation of services adapted to the needs of TVT centers

III. Software Architecture of the Proposed Solution

This paper aims to improve the quality of hybrid training, both theoretically and in terms of the development of tools that promote interaction and collaboration between the various actors in teaching and learning activities.

The architecture proposed (Figure 1) below consists of several software components interacting with each other via the network.

It aims to improve the quality of education in TVT centers through the following inputs.

A. Simultaneous Multi-communication Model

In our system, several services are outcome from WebRTC technology using VoIP to foster collaboration between educational participants. Gold, This technology is based on a "peer-to-peer" exchange mode allowing reduces the need for network infrastructure and minimizes latency without needing additional facilities.

These services hold into account socio-constructivist aspects that require the integration of several simultaneous connections.

Indeed, several researchers [18] advocate a mesh model based on a multipoint control unit (Figure 2) to centralize the processing of all audio and video streams. The MCU, also called a conference bridge, is a central gateway in a multipoint video conferencing system.

For this purpose, we used a software implementation of MCU based on a pairwise architecture named P2P-MCU to allow teachers to initiate simultaneous communications with their students via WebRTC.

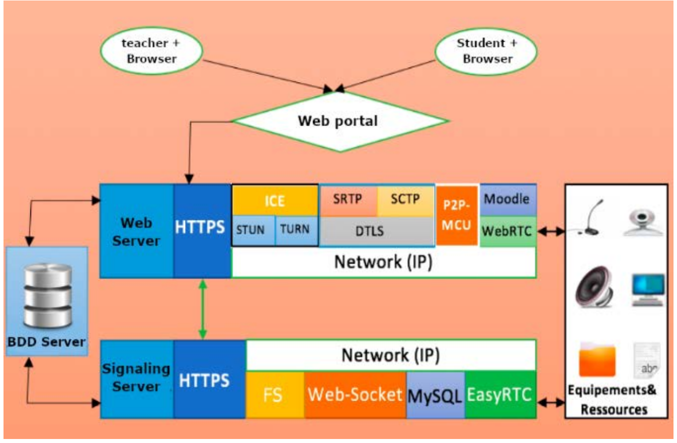


Fig. 1. Architecture Logicielle du Système D'enseignement.

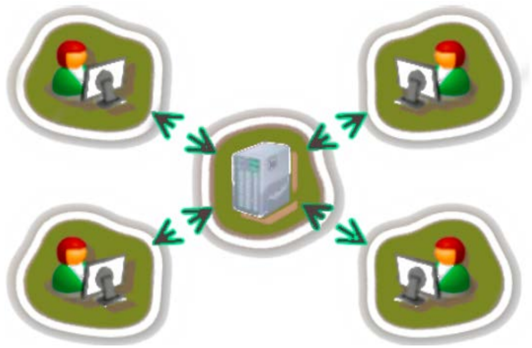


Fig. 2. "Mesh" Topology based on a Multipoint Control Unit.

B. Data Transit System

Most WebRTC capable devices sit behind one or more layers of NAT and may have security layers that block certain ports and protocols (sometimes with Deep Packet Inspection or DPI).

In addition, many of them placed behind corporate proxies and firewalls, not to mention firewalls and NAT on home Wi-Fi routers. In intranets, NAT gateways pose techniques that prevent direct communication between peers. Indeed, the main reason for this problem is because NAT corrects IP addresses and port numbers in order to hide private hosts.

As a result of the puncture technique limitation, it is necessary to use the services of an intermediate host that serves as a relay for the packets. This relay is usually located in the public Internet and relays packets in a direct communication between two hosts that are behind NAT.

For this specification, we use the relay tools STUN, TURN and ICE [19] to bypass NAT or firewall restrictions.

C. Media / Data Security Mechanism

Our scope of work requires real-time multimedia applications that require the use of SRTP as a transport protocol for the following reasons: fast delivery is preferred over reliable delivery and packet loss is acceptable to avoid delays resulting from rescheduling or retransmission of packets.

TLS (Transport Layer Security) cannot be used because it requires a reliable and slower protocol than a datagram-based protocol. In this perspective, DTLS (Datagram Transport Layer Security) brought adaptations to the TLS protocol to work in datagram communications by introducing counters and explicit messages [20].

Due to the unreliability of datagram-based protocols, DTLS incorporates mechanisms to retransmit the packet. This involves sending verification messages back to the server within a timeout period of 500-1000ms to ensure proper receipt of the messages. These verification messages also provide a means of avoiding denial of service (DoS) attacks, which in UDP, are very easy to carry out thanks to the exchange of a cookie.

In regard to these discussions, we have used in our scope the DTLS protocol as a basic security protocol for data and SRTP to ensure end-to-end confidentiality, message authentication and replay protection (Figure 3).

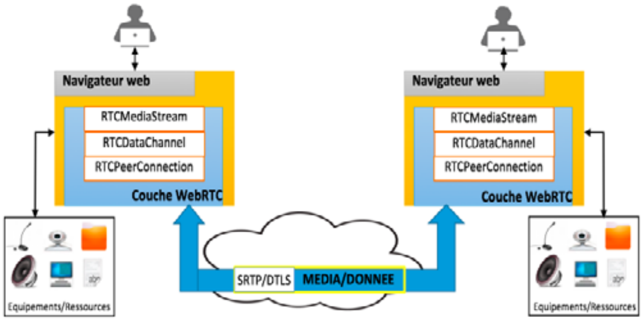


Fig. 3. Secure peer-to-peer Communication using DTLS/SRTP.

D. Signaling System

Signaling (Figure 4) is the mechanism for coordinating communication by exchanging identification and control messages between WebRTC pairs. In effect, it allows two or more WebRTC-capable web browsers to join, exchange contact information, negotiate a session that defines how they will communicate, and then finally establish the media channels between pairs for the transport of media streams exchanged directly between them [21].

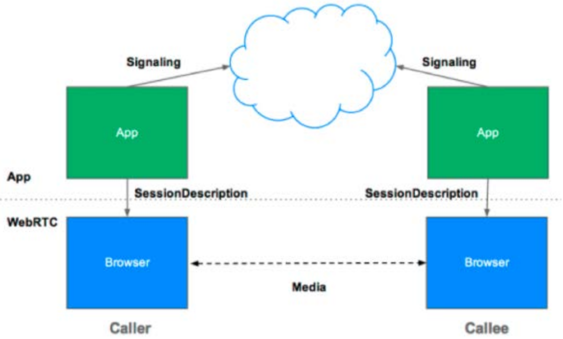


Fig. 4. Signaling Architecture.

The W3C and IETF standardization bodies have not imposed a particular signaling protocol in WebRTC in order to leave developers free to choose one of the existing protocols (SIP or Jingle) or to customize their own signaling protocol using web sockets. This signaling protocol choice strategy avoids redundancy and maximizes compatibility with already established technologies.

In this context, we used the EasyRTC implementation based on the Web-Socket protocol to create our own signaling system.

IV. Hybrid Learning Architecture

The proposed architecture (Figure 5) aims at proposing a device offering not only collaborative services while respecting the social-constructivist aspects but also the adaptation of the course flow to the learners' needs in order to ensure the same or better levels of interaction than those of traditional class's level.

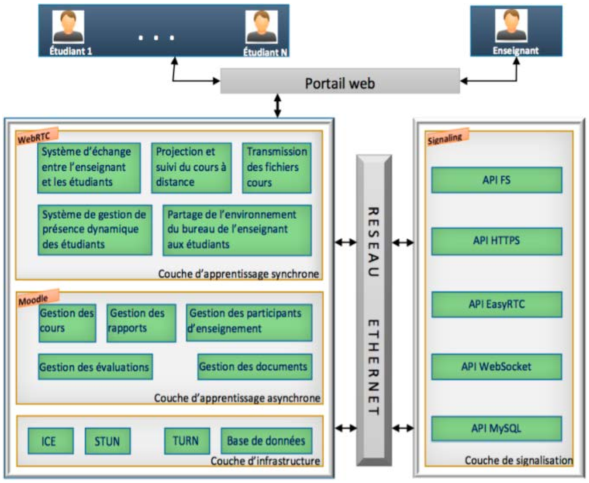

Étudiant 1
. . .
Étudiant N
Enseignant
Portail web
WebRTC
Système d'échange entre l'enseignant et les étudiants
Projection et suivi du cours à distance
Transmission des fichiers cours
Système de gestion de présence dynamique des étudiants
Partage de l'environnement du bureau de l'enseignant aux étudiants
Couche d'apprentissage synchrone
Moodle
Gestion des cours
Gestion des rapports
Gestion des participants d'enseignement
Gestion des évaluations
Gestion des documents
Couche d'apprentissage asynchrone
ICE
STUN
TURN
Base de données
Couche d'infrastructure
RESEAU ETHERNET
Signaling
API FS
API HTTPS
API EasyRTC
API WebSocket
API MySQL
Couche de signalisation

Fig. 5. Hybrid Learning Architecture.

Our approach allows TVT centers to improve the quality of their teaching system by integrating a hybrid system using both the capabilities of the Moodle platform and WebRTC technology. This architecture is structured in four layers:

1) Synchronous Learning Layer (SLL).
2) Asynchronous Learning Layer (ALL).
3) Infrastructures Layer (IL).
4) Signaling Layer (SL).

A. Synchronous Learning Layer (SLL)

This layer provides synchronous and real-time socio-constructivist learning activities based on WebRTC technology. Such a technology containing the MediaStream, PeerConnection and DataChannel APIs enables VoIP to operate in web browsers. Indeed, this layer essentially allows teachers and their students, through a simple browser:

1) To make projections and follow courses at a distance via video conferencing.

2) To transmit files (courses, exercises) to students in real time.

3) Share the teacher's office environment with students Interfaced with practical equipment.

4) To ensure the exchange system between the teacher and the students via instant messaging.

5) Disable the user's webcam.

6) Reactivate the previously deactivated webcam in the middle of a conversation.

7) Allow a teacher to eject a student from the system.

B. Asynchronous Learning Layer (ALL)

This layer, which ensures asynchronous learning, allows learners to follow the educational activities at their own pace. This flexibility of time for course delivery is a relevant solution to maximize access to higher education because it offers flexibility to the constraints that usually prevent professionals from participating in the classical course in universities.

To do so, we adopted the Moodle platform, which offers powerful and learner-centered tools. It also offers a collaborative environment that enhances both teaching and learning.

This layer also has a simple to use and easy to learn interface thanks to Moodle's constant usability improvements. It essentially offers the following tasks:

1) User, class and course management.

2) Scheduling of courses at a given date and time.

3) Course content management via a rich and versatile activity system.

4) Document management.

5) Reports and statistics management.

6) Learner evaluation management.

C. Signaling Layer (CL)

The signaling layer enables the creation of a real-time communication channel by defining a coordination mechanism between e-learning users.

Several developers have relied on the WebRTC API to provide libraries that hide the complexity associated with signaling and thus facilitate the development of WebRTC signaling servers. Figure 6 shows the software components of the signaling server. The operation of our system depends essentially on the signaling server implemented from the following modules:

1) Web-socket: allows creating bidirectional flows allowing the exchange in real time in both directions of communication.

2) EasyRTC: provides a library to simplify the development of WebRTC applications.

3) MySQL: allows connecting and making queries on the database.

4) FS: allowing the import of HTTPS server configuration files with SSL keys.

D. Infrastructures Layer (CS)

This layer hosts the physical servers allowing implement the physical infrastructure management applications. These are servers for traversing NAT/Firewall and the database server:

1) Servers to traverse NAT/Firewall: An ICE framework offered at this level, to overcome the difficulties of networking in the real world. ICE uses the highly reliable Xirsys STUN and TURN servers to try to find the best path to connect pairs. It tries all the possibilities in parallel and chooses the most efficient option that works. It tries to establish a connection using the host address from the operating system and a network card of a device.

2) Database servers: This layer contains a database server in order to keep or find all the data or information related to the educational activities of the system. The database is at the center of our computerized system allowing the collection, formatting, storage and use of data

V. Discussion and Result

A. Discussion

The proposed solution was used for one year by two technical and vocational training centers. These centers have experienced and appreciated the structural management of the trainings such as classes, pedagogical levels, teachers, students and annual enrollment in classes.

More than 40% of the students in these centers have benefited from online courses. They have been able to follow course sessions, practical work and even exams at a distance.

These students have shown remarkable progress in understanding and absorbing the courses, which is reflected in their exam scores. According to the observation, the online courses attract the intention of the students to cooperate and collaborate more with the teachers and each other to better understand the course concepts.

B. Results

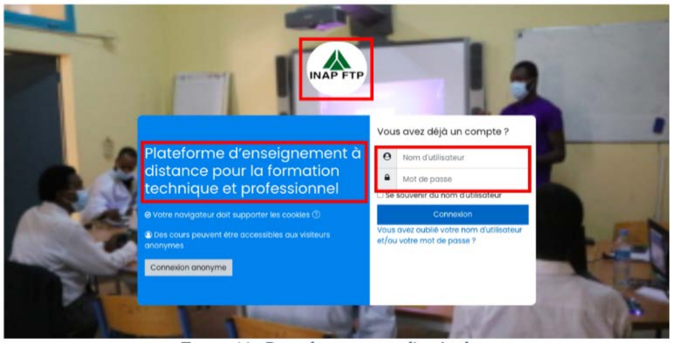
We designed the system following an architecture composed of two main parts such as the back-end and the front-end.

The back-end is the processing center that is at the heart of our distance learning system. It is a set of functionalities to be implemented in the INAP-FTP system in order to ensure the integration of the online course management in the core of this system.

The front-end is used to propose a space dedicated to the actual management of the course as described in our architecture. It is a system based on Moodle that allows us to propose practical mechanisms and standards for the management of quality pedagogical activities. Indeed, we take advantage of the set of relevant tools as proposed by Moodle without any intervention to modify the basic technical settings of Moodle components.

We then illustrate some system interfaces.

1) Login area: This space (Figure 6) allows users to authenticate via the above interface in order to be able to use the various distance-learning services.



Plateforme d'enseignement à distance pour la formation technique et professionnel

Votre navigateur doit supporter les cookies

Des cours peuvent être accessibles aux visiteurs anonymes

Vous avez déjà un compte ?

Nom d'utilisateur

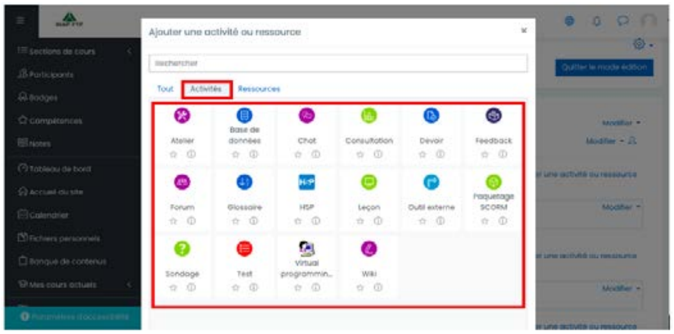
Mot de passe

□ Se souvenir du nom d'utilisateur

Vous avez oublié votre nom d'utilisateur et/ou votre mot de passe ?

Fig. 6. Authentification Page.

2) Asynchronous and interactive educational activities: The Figure 7 allows the teacher via the activities tab to add an activity to his course. Several activities offered in this case: Workshop, Homework, Forum, Lesson, Survey, Test and Chat.



Ajouter une activité ou ressource

Rechercher

Fig. 7. Asynchronous Activities Page.

3) Sharing the teacher's desktop environment with students: The Figure 8 allows the teacher via the activities tab to add an activity to his course. Several activities offered in this case: Workshop, Assignment, Forum, Lesson, Survey, Test and Chat.

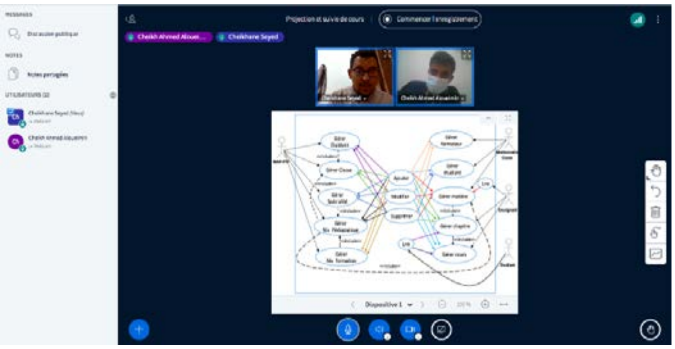


Fig. 8. Page for Teacher Desktop Screen Sharing.

VI. Conclusion

This paper deals with improving the quality of the TVT teaching system using innovative technologies adapted to the context of TVT centers in Mauritania.

Such a solution aims to offer relevant educational services both in synchronous and asynchronous mode between teachers and students.

However, innovation in distance education systems illustrates a concern for adaptation for those involved in education. It is more precisely about the integration of ICTs requiring configurations and adaptations for its members who often do not have sufficient experience in ICTs. This is why we have taken care in our work to choose flexible ICTs to use in order to concentrate as much as possible on the educational process.

The conclusive results of this work will enable education stakeholders to initiate and develop quality educational services through collaborative and innovative exchange systems. Our work provides evidence to dispel doubts about the quality of distance learning.

In terms of perspectives, our next steps will be to study the issues of availability and scalability in a context of optical fiber recently implemented in Mauritania.

References

[1] Yang Chen; KuangXinghong; Li Junjun; Wu Yanxiang, "A Blending E-Learning Model for Digital Electronic Technology Teaching," in E-Business and E-Government (ICEE), 2010 International Conference on , vol., no., pp.5355-5358, 7-9 May 2010.

[2] Anaraki, L.N.; Heidari, A., "Knowledge management process in digital age: Proposing a model for implementing e-learning through digital libraries," in Application of Information and Communication Technologies (AICT), 2011 5th International Conference on , vol., no., pp.1-5, 12-14 Oct. 2011.

[3] Y. Wu, M. Wen and C. Sun, "An architectural design teaching strategy based on online-to-offline (O2O) integration and blended learning," 2017 International Conference on Applied System Innovation (ICASI), 2017, pp. 111-113, doi: 10.1109/ICASI.2017.7988359.

[4] A. Oproescu, "The constructivist design of the assessment — an integrated teaching and learning process," *2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2017, pp. 1-4, doi: 10.1109/ECAI.2017.8166495.

[5] N. I. B. Adnan and Z. Tasir, "Online Social Learning Model," 2014 International Conference on Teaching and Learning in Computing and Engineering, 2014, pp. 143-144, doi: 10.1109/LaTiCE.2014.33.

[6] B. Lin, "The Research on Experience Teaching of Marketing Based on Constructivist Learning Perspective," 2009 First International Workshop on Education Technology and Computer Science, 2009, pp. 839-843, doi: 10.1109/ETCS.2009.724.

[7] S. Meng, F. Tao and L. Han, "The Joint Development of College Labor Education and Quality Education Based on the New Era," 2020 International Conference on Computers, Information Processing and Advanced Education (CIPAE), 2020, pp. 53-56, doi: 10.1109/CIPAE51077.2020.00021.

[8] Costa, JPV, Castro, MVL, Felcar, P., Mariano, AM et Souza, JCF (2021, juin). Facteurs de réussite dans les systèmes d'apprentissage en ligne pour les étudiants pendant la pandémie de COVID-19 : étude de cas dans un établissement d'enseignement supérieur brésilien. En 2021, 16e Conférence ibérique sur les systèmes et technologies de l'information (CISTI) (pp. 1-6). IEEE.

[9] Agarwal, A., Sharma, S., Kumar, V. et Kaur, M. (2021). Effet de l'apprentissage en ligne sur la santé publique et l'environnement pendant le confinement lié au COVID-19. Exploration et analyse de mégadonnées , 4 (2), 104-115.

[10] INAP-FTP., 2012 ; "Institut National de promotion de la formation technique et professtionnel", [Online], disponible sur : http://www.inap.mr/, accédé le 22 Mars 2022.

[11] Meyliana et al., "A Blockchain Technology-Based for University Teaching and Learning Processes," 2020 International Conference on Information Management and Technology (ICIMTech), 2020, pp. 244-247, doi: 10.1109/ICIMTech50083.2020.9211209.

[12] Y. Chen, "The Construction of Financial Teaching System Platform Based on Digital Intelligence Technology," 2021 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE), 2021, pp. 53-56, doi: 10.1109/ISAIEE55071.2021.00021.

[13] J. S. Mtebe and A. W. Kondoro, "Using Mobile Moodle to enhance Moodle LMS accessibility and usage at the University of Dar es Salaam," 2016 IST-Africa Week Conference, 2016, pp. 1-11, doi: 10.1109/ISTAFRICA.2016.7530649.

[14] Pedersen, J.M., Kuran, M.Ş. (2018). Moodle: Practical Advices for University Teachers. In: Choraś, M., Choraś, R. (eds) Image Processing and Communications Challenges 9. IP&C 2017. Advances in Intelligent Systems and Computing, vol 681. Springer, Cham. https://doi.org/10.1007/978-3-319-68720-9_21.

[15] H. Wan, L. Tang, Z. Zhong and Q. Cao, "Transit Traditional Face-to-Face Teaching to Online Teaching during the Outbreak of COVID-2019," 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 2020, pp. 355-362, doi: 10.1109/TALE48869.2020.9368330.

[16] S. Ouya, C. Seyed, A. B. Mbacke, G. Mendy and I. Niang, "WebRTC platform proposition as a support to the educational system of universities in a limited Internet connection context," 2015 5th World Congress on Information and Communication Technologies (WICT), 2015, pp. 47-52, doi: 10.1109/WICT.2015.7489643.

[17] S. Eltenahy, N. Fayez, M. Obayya and F. Khalifa, "Comparative Analysis of Resources Utilization in Some Open-Source Videoconferencing Applications based on WebRTC," 2021 International Telecommunications Conference (ITC-Egypt), 2021, pp. 1-4, doi: 10.1109/ITC-Egypt52936.2021.9513911.

[18] M. A. Hossain and J. I. Khan, "Dynamic MCU Placement for Video Conferencing on Peer-to-Peer Network," 2015 IEEE International Symposium on Multimedia (ISM), 2015, pp. 144-147, doi: 10.1109/ISM.2015.125.

[19] Dutton, S., 2013 ; "WebRTC in the real world : Stun, turn and signaling. Last accessed", [Online], disponible sur : http://www.html5rocks.com/en/tutorials/webrtc/infrastructure/, accédé le 22 Mars 2022.

[20] Daniele, T., 2012 ; Implementation and Evaluation of Datagram Transport Layer Security (DTLS) for the Android Operating System, Project Stockholm, Sweden June 2012.

[21] B. Sredojev, D. Samardzija and D. Posarac, "WebRTC technology overview and signaling solution design and implementation," 2015 38th International Convention on Information and Communication Technology.

Students' Characteristics of Student Model in Intelligent Programming Tutor for Learning Programming: A Systematic Literature Review

Rajermani Thinakaran[1]
Faculty of Data Science and Information Technology
INTI International University
Nilai, Negeri Sembilan, Malaysia

Suriayati Chuprat[2]
Razak Faculty of Technology and Informatics
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia

***Abstract*—This study describes preliminary results of a research related to Intelligent Programming Tutor (IPT) which is derived from Intelligent Tutoring System (ITS). The system architecture consists of four models. However, in this study student model mainly student characteristic was focused. From literature, 44 research articles were identified from a number of digital databases published between 1997 to 2022 base on systematic literature review (SLR) method. The findings show that the majority 48% of IPT implementation focuses on knowledge and skills. While 52% articles focused on a combination of two to three student characteristics where one of the combinations is knowledge and skill. When narrow down, 25% focused on knowledge and skills with errors or misconceptions; 4% focused on knowledge and skill with cognitive features; 5% focused focus on knowledge and skill with affective features; 2% focused on knowledge and skill with motivation; and 9% based on knowledge and skill with learning style and learning preferences as students' characteristics to build their student model. Whereas 5% focused on a combination of three student characters which are knowledge and skill with cognitive and affective features and 2% focused on knowledge and skill with learning styles and learning preferences and motivation as students' characteristics to construct the tutoring system student model. To provide an appropriate tutoring system for the students, students' characteristic needs to decide for the student model before developing the tutoring system. From the findings, it can say that knowledge and skills is an essential students' characteristic used to construct the tutoring system student model. Unfortunately, other students' characteristic is less considered especially students' motivation.**

***Keywords*—Intelligent tutoring system; intelligent programming tutor; student characteristics; student model**

I. Introduction

Intelligent Tutoring System (ITS) is a computer software system that can mimic the methods and dialog of natural human tutors, generate real time and on-demand instructional interactions as and when required by individual students. The implementation of ITSs also incorporate computational mechanisms and knowledge representations in the fields of Artificial Intelligence (AI) which addresses how to reason about intelligence together with multimedia and internet; Psychology on the other hand, consists of Cognitive Science which addresses how people think and learn, while the Education field focuses on how to provide the best support for teaching and learning [1] as illustrated in Fig. 1.

Fig. 2 depicts the evolution of ITS from the 1960s to the year 2000. The introduction of AI techniques and Expert Systems technology to CBI (Computer-Based Instruction) gave rise to ITS [2]. Early 2000, internet has become a central core to the educative environment, thus ITS incorporated with web platform so that the ITS can be accessed anytime and anywhere and known as Adaptive Web-Based Educational System (AWBES) [1].

An ITS architecture basically consists of four models [3] as illustrated in Fig. 3 which are 1) Domain Model - known as the expert or cognitive model. This model contains procedures, theories and problem-solving tactics of the domain to be learned; 2) Student Model - known as user or learner model. Considered as the main component of an ITS. The component gives special responsiveness to the students' cognitive and affective states and their progress in their learning process; 3) Tutoring Model - known as Pedagogical Model or Instructional Model. The model accepts information from the Domain Model and uses Student Model for making decisions on tutoring plans and actions; 4) Interface Model - provides the interface with which the students interact with the ITS.

The main aim of ITS is to improve students' learning process [4]. An ideal condition of the learning process is where students can receive lessons, resolve exercises and obtain immediate feedback. The feedbacks and hints are provided based on the analysis of the responses to each problem-solving step given by students.

In recent years, the development and improvement of ITS has been growing rapidly. Among some of the improvements include improvements on the problem-solving system that can support and help to give feedbacks and hints to students; improvements on model tracing that assesses students' current knowledge that facilitates the next step in order to support problem solving. In addition, improvements on knowledge tracing were also carried out that allows assessment of students' skills and knowledge level in order to release a new tutorial to facilitate learning and finally improvements on tutorial dialogues to support problem solving [5].

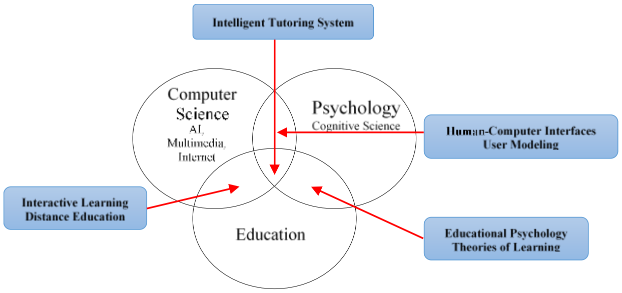


Fig. 1. The Development of ITS.

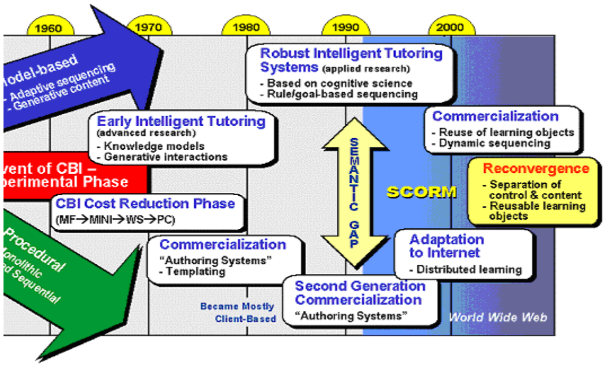


Fig. 2. The Evolution of ITS [2].

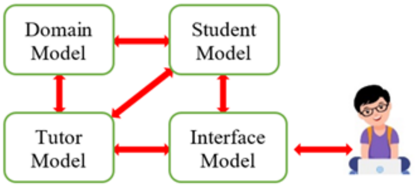


Fig. 3. ITS Architecture.

Educators agree that the most effective form of teaching is through one-to-one interaction with students [6]. In this context, ITS has an upper advantage as it provides personalized tutoring that is tailored to students' needs [5], [7], [8]. According to the findings from previous studies conducted by Kulik and Fletcher [9] and Colchester et al. [10], ITSs have successfully raised students' performance compared to those who were taught in conventional classes.

Chrysafiadi and Virvou [7] claim that, ITS to be become more adaptive and personalize, students' characteristic as student model need to be considered. The students' characteristic comprises of knowledge and skill; errors or misconceptions; learning styles and learning preferences; cognitive features; motivation; and affective features.

Knowledge refers to familiarity with theoretical concepts and factual information and skills refer to the proficiencies developed through practice [7]. During the learning process, errors or misconceptions can be identified. The concept of error or misconceptions can be defined as a process or fact that does not match a given norm [11]. Learning Styles and Learning Preferences - refer to how a student identifies, gathers and processes their learning materials [12]. While Cognitive features refer to students' aspects such as attention, knowledge, ability to learn and recall memory, opinion, attention, collaborative skills, capabilities to solve problems and make conclusions, analyzing abilities and critical thinking [7]. Literally motivation is the desire to do things. Motivation plays a significant role in students' learning process [13]. Emotional factors are known as affective features such as sadness, happiness, frustrations, anger, interest, boredom, distractions, aims and confusion [14]. Subsequently, affective features can be based on students' motivations [7].

Among the stated students' characteristics, motivation is considered as the main factor for engaging students in their learning [15], [16] and in academic performance [17]. Meanwhile, Abuhmaid [18], Hamzah et al., [19] and Sundar and Kumar [20] argue that students' motivation is an important factor in ensuring the success of ITS implementation. Abuhmaid [18] also pointed that motivation factors need to be considered when designing any ITS materials. Studies on the relationship between motivating factors and learning have been a prominent research topic in the field of education as well as studies focusing on eLearning [15].

McGill [21], for example, studies the use of robots to influence students' motivation when learning introductory programming. In order to ensure students feel motivated to use eLearning, Hamzah et al., [19] applied the ARCS+G (Attention, Relevance, Confidence, Satisfaction + Gamification) as motivational design model in the development process in their study. Another study conducted by Nikou and Economides [22] examined the impact of using mobile devices during the learning activity on students' learning motivation. Results obtained by Abuhmaid [18] reveal that utilizing the flipped learning strategy in an eLearning environment has a significant improvement on students' motivation to learn. Tambunan, Rusdi and Miarsyah [23] suggest that a combined usage of eLearning with a Problem Based Learning (PBL) model and motivation is an effective way to improve students' learning outcomes. Even though some may argue that a game environment can be used to ensure a good transfer of knowledge in a fun way, Yedri et al., [24] claim that a balance between learning transfer and motivation is the major key to success.

A. Intelligent Tutoring System for Learning Programming

Programming tools have been actively researched in their effectiveness to support teaching and learning. Pears and his colleagues [25] have summarized these programming tools into five categories one of which is ITSs. As highlighted in the previous section, ITSs provide many benefits in students' learning process.

An IPT (Intelligent Programming Tutor) is a specific implementation of an ITS for learning programming. The ideas behind the use of IPT are to create a learning process where students can receive tutelage, resolve exercises and receive instant feedbacks imitating one-to-one human tutoring.

As explained above, IPTs is derived from ITS' ideas in which students' characteristics also need to be considered in creating a conducive and effective learning process. In the following section, an exhaustive systematic literature review (SLR) was carried out to identify what types of student characteristics were used to design the IPT.

II. Method

In this section, a SLR (Systematic Literature Review) was carried out to obtain answers to the following question: What types of student characteristics were used to design the IPT? SLR was conducted in this study as it is a process that can be used for recognizing, evaluating and interpreting research materials to answer several research questions [26].

To answer the question stated above, PICOC as proposed by Petticrew and Robert [27] was used in the study. PICOC comprise of five elements which are Population, Intervention, Comparison, Outcomes and Context. Table I shows a summary of PICOC for this study.

TABLE I. Summary of PICOC

Population	Student
Intervention & Comparison	Intelligent Programming Tutor or Intelligent Tutoring System
Outcomes	Student characteristics in Intelligent Programming Tutor or Intelligent Tutoring System
Context	Reviews of all studies of Intelligent Programming Tutor or Intelligent Tutoring System within the domain of Programming subject

The identification of primary sources from journals, conferences and online databases is important to ensure a wide coverage of potential sources. A survey of literature included all research works published from online database such as ACM Digital Library, Google Scholar, IEEE Explore, ISI Web of Knowledge, Science Direct and Springer. These online databases were selected to be used in this study as they are the most popular and frequent databases used by previous researchers in investigating the use of eLearning. In addition, references retrieved from SLR articles were analyzed to identify any literature that may have been ignored or overlooked during the search.

Using Booleans of AND, OR in the keyword combinations conducted include keywords such as intelligent programming tutor; intelligent tutoring system AND programming; and programming AND intelligent tutoring system. The use of the Boolean OR is to incorporate alternative synonyms and spellings while the usage of the Boolean AND is to link the major terms.

A search of these databases and journals allow the data collection to be inclusive and comprehensive. A total of 73 papers were selected which the searching technique described above. These articles were reviewed while papers that were not categorized under any refereed journal articles such as proceedings or editor-reviewed papers were excluded from the final analysis. The focus of this review was on refereed articles which assist in ensuring the quality and relative rigor of data sources. Therefore, research papers that did not conduct an in-depth discussion of their ITS in programming particularly those that did not investigate student characteristic(s) were omitted. Papers that were not published in English language and gray papers such as those without any bibliographic information (publication date/type, volume and issue numbers) were also excluded. Duplicated papers were also excluded (only the most recent, complete and improved one is included) from the SLR in this study. In total, 44 relevant papers which were published between 1997 and 2022 were gathered and thoroughly examined in this study.

III. Result and Discussion

A. Data Analysis

In Table II, the 44 articles were organized based on what type of student characteristics was used to construct the user model; what modelling technique was applied to construct the user model for the intended IPT system; the subject domain and learning environment.

B. Synthesis of SLR on Student Characteristics for Student Model in IPTs

Fig. 4 illustrates how students' characteristics were considered in constructing user model mainly in IPTs base on 44 articles which was revealed in Table II.

From Fig. 4, 48% or 21 articles were found mainly focused on knowledge and skills as student characteristic for their user model. The objective of this characteristic is to improvised students' theoretical concepts knowledge and programming proficiencies skills in particular topic of programming subject. To develop the user model base on knowledge and skills as student characteristic, different researchers use different modelling technique such as Bayesian network, Markov Decision Process, Regression model, Rule Base and K* classifier. However, the common modelling technique is Rule Base because it is easier to build due to improved authoring tools and remain a popular option.

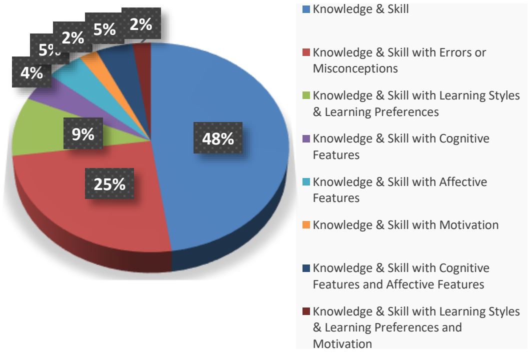


Fig. 4. Students' Characteristics for user Model in IPTs.

TABLE II. SUMMARY OF IPT REVIEW

Student Characteristic	IPT	Modelling Technique	Subject Domain and Learning Environment
Knowledge & Skill	ALLIGATOR [28]	-	To teach data flow diagram using visual programming environment with multiple informative and tutoring feedback components.
	AOMS [29]	-	To teach graph using C Programming
	BITS [30]	Bayesian network	To teach programming concept using C++.
	BOTS [31]	Markov Decision Process	To teach pseudocode in using game environment.
	ChiQat – Tutor [32]	Regression model	Using visualization to teach link list.
	COLLEGE [33]	-	Editing, compiling the source code, and executing the object code with aid of animation and visualization.
	CIMEL ITS [34]	Rule Base	To teach OOP concepts and observe students' progress and offer assistance based on pedagogical strategies adapted to the individual student.
	CPP-Tutor [35]	-	To teach programming concept using C++ and provide hints and feedback during problem-solving in tutorials.
	CS-I [36]	Rule Base	To teach programming concept using C++ and detect each students' level of understanding.
	DrJava [37]	-	To write, test and debug Java programs.
	EduJudge [38]	-	Submission, management and automatic evaluation of programming exercises.
	KSC-PaL [39]	K^* classifier	Collaborate with students to solve problems on data structures mainly on linked lists, stacks, and binary search tree.
	Marmoset [40]	-	To write and test Java code and helps the instructor to monitor student progress.
	OmniCode [41]	-	To teach novice students basic programming concept using Python.
	ProgTool [42]	-	Using visualization to teach OOP concepts.
	PASS [43]	-	Assisting beginners in learning programming.
	PLTutor [44]	Rule Base	To teach syntax and semantic of JavaScript using visualization.
	PLWeb [45]	-	Assist instructors to design computer programming exercises and to help students to study and practice programming exercises.
	SCALE [46]	-	Teach multiple topics consisting of pseudocode, sequential search, binary search subprograms and recursive.
	ViLLe [47]	-	Developed to visualize programming syntax written by students in Java or C++.
	WebTask [48]	-	To write Java code in a method body, testing, and received feedback in animation and visualization environment.
Knowledge & Skill with Errors or Misconceptions	ADIS [49]	Constraint Based Modelling	To teach basic algorithms of linked-lists, stacks, queues, trees and graph by visually.
	ADIL [50]	-	To teach C language and explain logical errors.
	AutoLEP [51]	-	Helps students to find and work through bugs in C language and also provides immediate detailed feedback.
	Collab ChiQat [52]	-	Developed to teach linked lists, stacks, and binary search trees in collaboration environment.
	iList [53]	-	Helps students to learn linked lists in visualization form.
	iSnap [54]	Contextual Tree Decomposition algorithm	To teach programming control structure using block-based programming.
	INCOM [55]	Constraint Based Modelling	Help students on programming logic using Prolog.
	J-LATTE [56]	Constraint Based Modelling	To teach Java in terms of design and syntax.
	OOPs [57]	Constraint Based Modelling	Help students to understand and overcome their misconceptions in OOP and reinforce the correct learning methods.
	ProBot [58]	Rule Base	Use game concept to improve students' abilities in programming control structures.

	@KU-UZEM [59]	Constraint Based Modelling	To teach C language and to overcome misconceptions in terms of concept and syntax.
Knowledge & Skill with Learning Styles & Learning Preferences	ABITS [60]	Association Rule Mining And Fuzzy C-Means Clustering	Introduction to Java Programming.
	EDUCA [61]	Neural Networks	Introduction to Maya Programming Language.
	ELM-ART [62]	-	Introduction to LISP Programing. Sample based problem solving support, detailed analysis of student answers, solving support, reminder option.
	Protus [63]	-	Help students during programming learning process by advising students to take an appropriate action when needed, monitoring their progress and tracking student learning styles.
Knowledge & Skill with Cognitive Features	APT [64]	ACT-R theory	To write short programs in Lisp, Pascal or Prolog
	WPAS [65]	-	Supporting programming learning activities with various difficulty levels.
Knowledge & Skill with Affective Features	E-Learning 3.0 [66]	Fuzzy-Logic and Nayve Bayes classifier algorithm	To teach Java according student emotions.
	PIT [67]	-	Teach programming skill and provide feedback based on students' emotion.
Knowledge & Skill with Motivation	FITS [68]	Bayesian networks	To teach flowchart using game environment.
Knowledge & Skill with Cognitive Features and Affective Features	Java Sensei [69]	Neural Networks, Fuzzy Logic	To teach Java and analyze student cognitive and emotion level during using the system.
	JavaTutor [70]	ACT-R theory & machine learning techniques	Teach Java by interacting human-to-computer and body expressions.
Knowledge & Skill with Learning Styles & Learning Preferences and Motivation	LOs [71]	Rule Base	Used simulation-based to teach array sorting.

Whereas another 52% or 23 articles focused on a combination of two to three student characteristics where one of the combinations is knowledge and skill. From the study, it was identified that 25% or 11 articles focused on knowledge and skills with errors or misconceptions as students' characteristics to build the student model. The researchers aim is to improvise the students' knowledge by learning from mistakes. From the educational point of view, learning from mistake or error can be powerful learning process, especially for learning programming. Students learn much faster when they made mistake first, especially in programming. In other words, getting the incorrect answer helps them to remember the correct one. To develop the student model base on these two students' characteristics, the researchers has considered three different modelling techniques which are Constraint Based Modelling, Contextual Tree Decomposition algorithm and Rule Base. Among these three techniques, Constraint Based Modelling is most preferred by the researcher because the algorithm was originally developed as a hypothesis about how student learn from their mistakes.

Two articles or 4% focused on knowledge and skill with cognitive features as students' characteristics for their student model. The cognitive features are students' ability to learn and recall memory and also capabilities to solve problems and make conclusion were used to construct the user model. ACT-R (Adaptive Control of Thought—Rational) theory was used to develop the tutoring system student model. The theory using a cognitive architecture that uses production rules to model student problem solving processes.

Another two articles or 5% focused focus on knowledge and skill with affective features as students' characteristics for their student model. This model was able to recognize and analyze student emotion such as frustration, boredom, engagement, confusion and excitement through students' facial expressions. The system was developed base on Fuzzy-Logic and Naive Bayes classifier algorithm.

One article or 2% focused on knowledge and skill with motivation as students' characteristics to construct the user model. The model was developed using Bayesian networks in game environment call tic-tac-toe. While another four article or 9% construct the student model based on knowledge and skill with learning style and learning preferences. The model able to track students' learning styles during their learning process by advising students to take an appropriate action when needed. Association Rule Mining and Fuzzy C-Means Clustering and Neural Networks were considered as modelling technique to construct the student model base on students' characteristics.

There are two articles or 5% focused on a combination of three student characters which are knowledge and skill with cognitive and affective features. The user model was developed to analyze students' cognitive and emotional conditions during the learning process. The model use ACT-R theory for knowledge representation while affective features were obtained through body expressions using sensor detection. The detection performed using machine learning techniques.

Lastly, one article (2%) focused on knowledge and skill with learning styles and learning preferences and motivation as students' characteristics to construct the tutoring system student model. The model able to detect students' learning styles and preferences using some predefine rules during the learning process and to motivate the students, simulation and visualization was used in the system user interface design.

From Jamal and Naemah [72] point of view, the effective teaching of programming subjects can be achieved by providing an appropriate tutoring system for the students. To achieve this, what type of students' characteristic need to be decided for the student model before developing the tutoring system [7]. From Fig. 4, it can say that knowledge and skills is an essential students' characteristic used to construct the tutoring system student model. Unfortunately, other students' characteristic is less considered especially students' motivation.

From the findings presented in Table II and Fig. 4, it can be seen that only 4.0% or 2 articles [68], [71 considered motivation as a student characteristic for the student model in IPT. On the other hand, Hooshyar et al. [68] used the game approach to motivate students to learn programming algorithm while Tuparov, Tuparov and Jordanov [71] used simulation-based ITP to help motivate students to understand array sorting. These motivations only encourage students as per view only.

Based on the results obtained from the existing literature thoroughly discussed above, it can be concluded that there is a lack of focus on motivation as a students' characteristic for student model mainly in IPT and generally in ITSs. Since student motivation is an important factor [15], [73] in learning programming [74], therefore the same consideration needs to be considered at the IPTs level and also ITSs.

IV. Conclusion

IPT is derived from ITSs. To develop an IPT system, students' characteristics need to be considered first before construct the student model which is one of important model in ITS architecture. From this study, it was identified that motivation was less considered as students' characteristics in constructing student model for IPT and generally in ITSs. Motivation and learning are highly complex aspects of human behaviour. Motivation has been agreed as a crucial aspect affecting learning behaviour, learning process and learning achievement. So, the same concern need to be consider in tutoring system implementation where can bring numerous benefits.

References

[1] A. Alkhatlan and J. Kalita, "Intelligent tutoring systems: A comprehensive historical survey with recent developments," International Journal of Computer Applications (0975 - 8887), Volume 181 - No.43, March 2019.

[2] L. Samuelis, "Notes on the components for intelligent tutoring systems," Acta Polytechnica Hungarica, 4(2), pp. 77-85, 2007.

[3] A. K. Erümit and İ. Çetin, "Design framework of adaptive intelligent tutoring systems," Education and Information Technologies, 25(5), pp. 4477-4500, 2020.

[4] B. Vesin, M. Ivanović, A. Klašnja-Milićević, and Z. Budimac, "Personal Assistance Agent in Programming Tutoring System", In Agent and Multi-Agent Systems: Technologies and Applications, pp. 441-451, Springer International Publishing, 2015.

[5] E. Dehkourdy, A. Reza, R. Mohasanati, and S. Hakimnia, "Main Components of Intelligent Tutoring Systems", Life Science Journal, 10(8), 2013.

[6] H. Bui, "A Classification of Data-Driven Hint Generation Techniques for Code-Writing Intelligent Tutoring Systems," American Journal of Computer Science and Information Engineering, 4(2), pp. 16-23, 2017.

[7] K. Chrysafiadi, and M. Virvou, "Student modeling approaches: A literature review for the last decade', *Expert Systems with Applications*, 40(11), pp. 4715-4729, 2013.

[8] T. W. Price, Y. Dong, and D. Lipovac, "iSnap: Towards Intelligent Tutoring in Novice Programming Environments," In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, pp. 483-488, ACM, March 2017.

[9] J. A. Kulik, and J. D. Fletcher, J. D. "Effectiveness of intelligent tutoring systems: a meta-analytic review', Review of Educational Research, 86(1), pp. 42-78, 2016.

[10] K. Colchester, H. Hagras, D. Alghazzawi, and G. Aldabbagh, "A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms," Journal of Artificial Intelligence and Soft Computing Research, 7(1), pp. 47-64, 2017.

[11] J. Ljubomir, "Teaching Introductory Programming: Agent-based Approach with Pedagogical Patterns for Learning by Mistake," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No.6, 2014.

[12] M. A. Ghazal, N. A. M. Zin, and Z. Muda, "Designing Domain Model For Adaptive Web-based Educational System According to Herrmann Whole Brain Model," Journal of Engineering Research and Technology, 3(3), 2016.

[13] T. Khan, K. Johnston, and J. Ophoff, "The Impact of an Augmented Reality Application on Learning Motivation of Students," Advances in Human-Computer Interaction, Volume 2019, 2019.

[14] C. Cunha-Pérez, M. Arevalillo-Herráez, L. Marco-Giménez, and D. Arnau, "On Incorporating Affective Support to an Intelligent Tutoring System: an Empirical Study," IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 13(2), pp.63-69, 2018.

[15] Y. Azliza, A. Noraida, Y. M. Hafiz, M.S.M. Yazid Sukinah and S. N. Suhana, "Learning Motivation Assessment Model: A Review," Australian Journal of Basic and Applied Sciences, 8(4) Special 2014, pp. 163-169, 2014.

[16] F. T. Leow, and M. Neo, "Peer Interaction and Students' Perceptions Towards Constructivist-Collaborative Learning Environment: Motivation and Affective Factor," In ICEL2016-Proceedings of the 11th International Conference on e-Learning: ICEl2016, pp. 87, Academic Conferences and publishing limited, June 2016.

[17] J. Cibulka, and G. A. Giannoumis, "Augmented and Virtual Reality for Engineering Education," In Proceedings of the 58th Conference on Simulation and Modelling (SIMS 58), No. 138, pp. 209-219, Linköping University Electronic Press, Sptember 2017.

[18] A. Abuhmaid, "The Impact of Using Flipped Learning Strategy on Students' motivation for Learning," 10th annual International Conference of Education, Research and Innovation, Seville (Spain), 2017.

[19] W. M. A. F. W. Hamzah, N. H. Ali, M. Y. M. Saman, M. H. Yusoff, and A. Yacob, "Influence of gamification on students' motivation in using e-learning applications based on the motivational design model," International Journal of Emerging Technologies in Learning (iJET), 10(2), pp. 30-34, 2015.

[20] P.P. Sundar, and A. S. Kumar, "A systematic approach to identify unmotivated learners in online learning," Indian Journal of Science and Technology, 9(14), 2016.

[21] M. M. McGill, "Learning to program with personal robots: Influences on student motivation," ACM Transactions on Computing Education (TOCE), 12(1), pp. 4, 2012.

[22] S. A. Nikou, and A. A. Economides, (2016) "The impact of paper-based, computer-based and mobile-based self-assessment on students' science motivation and achievement," Computers in Human Behavior, 55, pp. 1241-1248, 2016.

[23] L. Tambunan, R. Rusdi, and M. Miarsyah, "Efectiveness of Problem Based Learning Models by Using E-Learning and Learning Motivation

Toward Students Learning Outcomes on Subject Circullation Systems," Indonesian Journal of Science and Education, 2(1), pp. 96-104, 2018.

[24] O. B. Yedri, L. El Aachak, A. Belahbib, H. Zili, and M. Bouhorma, "Motivation Analysis Process as Service Applied on Serious Games," International Journal of Information Science and Technology, 2(1), pp. 4-11, 2018.

[25] A. Pears, S. Seidman, C. Eney, P. Kinnunen, and L. Malmi, "Constructing a core literature for computing education research," ACM SIGCSE Bulletin, 37(4), pp. 152-161, 2005.

[26] R. Thinakaran, and R. Ali, "Work in progress: An initial review in programming tutoring tools," In 2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), pp. 1-4, IEEE, December 2015.

[27] M. Petticrew, and H. Roberts, Systematic reviews in the social sciences: A practical guide. John Wiley & Sons, 2008.

[28] Mosconi, M., Ottelli, D. and Porta, M. (2003) 'Alligator, a Web-based Distributed Visual Programming Environment', WWW (Posters), pp. 3.

[29] M. Gaeta, F. Orciuoli, and P. Ritrovato, "Advanced ontology management system for personalised e-Learning," Knowledge-Based Systems, 22(4), pp. 292-301, 2009.

[30] C. J. Butz, S. Hua, and R. B Maguire, "A web-based intelligent tutoring system for computer programming," In Web Intelligence, 2004. WI 2004. Proceedings, IEEE/WIC/ACM International Conference on pp. 159-165, IEEE, September 2004.

[31] A. Hicks, Y. Dong, R. Zhi, V. Cateté, and T. Barnes, "BOTS: Selecting Next-Steps from Player Traces in a Puzzle Game," In EDM (Workshops), June 2015.

[32] N. Green, B. Di Eugenio, R. Harsley, D. Fossati, O. AlZoubi, and M. Alizadeh, "Student behavior with worked-out examples in a computer science intelligent tutoring system," In International Conference on Educational Technologies, November 2015.

[33] C. Bravo, M. J. Marcelino, A. J Gomes, M. Esteves, and A. J. Mendes, "Integrating Educational Tools for Collaborative Computer Programming Learning," J. UCS, 11(9), pp. 1505-1517, 2005.

[34] S. H. Moritz, F. Wei, S. M. Parvez, and G.D. Blank, "From objects-first to design-first with multimedia and intelligent tutoring," In ACM SIGCSE Bulletin, Vol. 37, No. 3, pp. 99-103, ACM, June 2005.

[35] S. Naser, "Evaluating the effectiveness of the CPP-Tutor an intelligent tutoring system for students learning to program in C++," Journal of Applied Sciences Research, 5(1), pp. 109-114, 2009.

[36] J. P. Yoo, S. J. Seo, and S. K. Yoo, "Designing an Adaptive Tutor for CS-I Laboratory," In International Conference on Internet Computing, pp. 459, 2004.

[37] E. Allen, R. Cartwright, and B. Stoler, "DrJava: A lightweight pedagogic environment for Java," In *ACM SIGCSE Bulletin*, Vol. 34, No. 1, pp. 137-141, ACM, February 2002.

[38] E. Verdú, L. M. Regueras, M. J. Verdú, J. P. Leal, J. P. de Castro, and R. Queirós, "A distributed system for learning programming on-line," Computers & Education, 58(1), pp. 1-10, 2012.

[39] C. Howard, P. Jordan, B. Di Eugenio, and S. Katz, S. "Shifting the load: A peer dialogue agent that encourages its human collaborator to contribute more to problem solving," International Journal of Artificial Intelligence in Education, 27(1), pp. 101-129, 2017.

[40] J. Spacco, D. Hovemeyer, W. Pugh, F. Emad, J. K. Hollingsworth, and N. Padua-Perez, "Experiences with marmoset: designing and using an advanced submission and testing system for programming courses," ACM Sigcse Bulletin, 38(3), pp. 13-17, 2006.

[41] H. Kang, and P. J. Guo, "Omnicode: A Novice-Oriented Live Programming Environment with Always-On Run-Time Value Visualizations,' 2017.

[42] M. Goyal, "Development of agent-based intelligent tutoring system for teaching object-oriented programming concepts," In Proceedings of the 9th International Conference on Education and Information Systems, Technologies and Applications (EISTA 2011), pp. 17-22, 2011.

[43] K. M. Law, V. C. Lee, and Y. T. Yu, "Learning motivation in e-learning facilitated computer programming courses," Computers & Education, 55(1), pp. 218-228, 2010.

[44] G. L. Nelson, B. Xie, and A. J. Ko, "Comprehension First: Evaluating a Novel Pedagogy and Tutoring System for Program Tracing in CS1," In Proceedings of the 2017 ACM Conference on International Computing Education Research, pp. 2-11, ACM, August 2017.

[45] S. H. Tung, T. T. Lin, and Y. H. Lin, "An exercise management system for teaching programming," Journal of Software, 8(7), pp. 1718-1725, 2013.

[46] I. Verginis, A. Gogoulou, E. Gouli, M. Boubouka, and M. Grigoriadou "Enhancing learning in introductory computer science courses through SCALE: An empirical study", IEEE transactions on education, 54(1), pp. 1-13, 2011.

[47] T. Rajala, M. J. Laakso, E. Kaila, and T. Salakoski, T. "Effectiveness of Program Visualization: A Case Study with the ViLLE Tool," Journal of Information Technology Education, 7, 2008.

[48] G. Rößling, "A family of tools for supporting the learning of programming," Algorithms 2010, 3(2), pp.168-182, 2010.

[49] K. Warendorf, and C. Tan, "ADIS-An animated data structure intelligent tutoring system or Putting an interactive tutor on the WWW," In Proceedings of Workshop Intelligent Educational Systems on the World Wide Web at AI-ED, Vol. 97, pp. 54-60, August 1997.

[50] A.M. Zin, S. A. Aljunid, Z. Shukur, and M. J. Nordin, "A Knowledge-based automated debugger in learning system," Proceedings of the 4th International Workshop on Automated Debugging, (WAD' 01), ACM Press, Munich, 2001.

[51] T. Wang, X. Su, P. Ma, Y. Wang, and K. Wang, "Ability-training-oriented automated assessment in introductory programming course," Computers & Education, 56(1), pp. 220-226, 2011.

[52] R. Harsley, D. Fossati, B. Di Eugenio, and N. Green, "Interactions of Individual and Pair Programmers with an Intelligent Tutoring System for Computer Science," In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, pp. 285-290, ACM, March 2017.

[53] D. Fossati, B. Di Eugenio, S. Ohlsson, C. Brown, and L. Chen, "Data driven automatic feedback generation in the iList intelligent tutoring system," Technology, Instruction, Cognition and Learning, 10(1), pp. 5-26, 2015.

[54] T. W. Price, Y. Dong, and D. Lipovac, "iSnap: Towards Intelligent Tutoring in Novice Programming Environments," In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, pp. 483-488, ACM, March 2017.

[55] N. T. Le, W. Menzel, and N. Pinkwart, "Evaluation of a Constraint-Based Homework Assistance System for Logic Programming," In Proceedings of the 17th International Conference on Computers in Education, 2009.

[56] J. Holland, A. Mitrovic, and B. Martin, "J-LATTE: A Constraint-based Tutor for Java," Proceedings of the 17th International Conference on Computers in Education, ICCE 2009, pp. 142–146. Asia-Pacific Society for Computers in Education, Hong Kong, 2009.

[57] J. Gálvez, E. Guzmán, and R. Conejo, R. (2009) "A blended E-learning experience in a course of object oriented programming fundamentals," Knowledge-Based Systems, 22(4), pp. 279-286, 2009.

[58] J. Moreno, "Digital Competition Game to Improve Programming Skills," Journal of Educational Technology & Society, 15(3), pp. 288, 2012.

[59] U. Kose, and O. Deperlioglu, "Intelligent learning environments within blended learning for ensuring effective c programming course", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.1, pp. 105 – 124, 2012.

[60] T. T. Sampathkumar, R. Gowri, and V. Venkateswaran, "Designing an adaptive distributed tutoring system based on Students' learning style and collaborative learning using intelligent agents. International Journal of Computer Applications, 87(17), 2014.

[61] R. Z. Cabada, M. L. B. Estrada, and C. A. R. García, "EDUCA: A web 2.0 authoring tool for developing adaptive and intelligent tutoring systems using a Kohonen network," Expert Systems with Applications, 38(8), pp. 9522-9529, 2011.

[62] G. Weber, and P. Brusilovsky, "ELM-ART–An interactive and intelligent web-based electronic textbook," International Journal of Artificial Intelligence in Education, 26(1), pp. 72-81, 2016.

[63] B. Vesin, M. Ivanović, A. Klašnja-Milićević, and Z. Budimac, "Personal Assistance Agent in Programming Tutoring System," In Agent and Multi-Agent Systems: Technologies and Applications, pp. 441-451, Springer International Publishing, 2015.

[64] A. Corbett, "Cognitive mastery learning in the ACT programming tutor," AAAI Technical Report SS-00-01, 2000.

[65] W. Y. Hwang, R. Shadiev, C. Y. Wang, and Z. H. Huang, "A pilot study of cooperative programming learning behavior and its relationship with students' learning performance," Computers & Education, 58(4), pp. 1267-1281, 2012.

[66] R. Z. Cabada, M. L. B.cEstrada, F. G.cHernández, R. O. Bustillos, and C. A. Reyes-García, "An affective and Web 3.0-based learning environment for a programming language," Telematics and Informatics, 2017.

[67] Tiam-Lee, T. J. and Sumi, K. (2018, June) Adaptive Feedback Based on Student Emotion in a System for Programming Practice. In International Conference on Intelligent Tutoring Systems, pp. 243-255, Springer, Cham., June 2018.

[68] D. Hooshyar, R. B. Ahmad, M. Yousefi, M. Fathi, S. J. Horng, and H. Lim, "Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills," Computers & Education, 94, pp.18-36., 2016.

[69] M. L. Barrón-Estrada, R. Zatarain-Cabada, F. G., Hernández, R. O. Bustillos, and C. A. Reyes-García, "An Affective and Cognitive Tutoring System for Learning Programming," In Mexican International Conference on Artificial Intelligence, pp. 171-182, Springer, Cham, October 2015.

[70] J. B. Wiggins, K. E. Boyer, A. Baikadi, A. Ezen-Can, J. F. Grafsgaard, E. Y. Ha, and E. N. Wiebe, "JavaTutor: an intelligent tutoring system that adapts to cognitive and affective states during computer programming," In Proceedings of the 46th ACM Technical Symposium on Computer Science Education, pp. 599-599, ACM, February 2015.

[71] G. Tuparov, D. Tuparova, and V. Jordanov, "Teaching sorting and searching algorithms through simulation-based learning objects in an introductory programming course," Procedia-Social and Behavioral Sciences, vol. 116, pp. 2962-2966, 2014.

[72] O. Jamal, and A. Naemah, "The Uncommon Approaches of Teaching the Programming Courses: The Perspective of Experienced Lecturers," Computing Research & Innovation (CRINN), Vol. 1, pp. 64, 2016.

[73] A. de Vicente, Towards Tutoring Systems that Detect Students' Motivation: An Investigation. PhD. thesis, Institute for Communicating and Collaborative Systems, School of Informatics, University of Edinburgh, 2003.

[74] M. M. McGill, "Learning to program with personal robots: Influences on student motivation," ACM Transactions on Computing Education (TOCE), 12(1), pp. 4, 2012.

News Analytics for Business Sentiment Suggestion

Sirinda Palahan
School of Science and Technology
University of the Thai Chamber of Commerce
Bangkok, Thailand

***Abstract*—Business and economics news has become one of the factors businesses consider when making decisions. However, the exponential increase in the availability of business information sources on the internet makes it more difficult for entrepreneurs to keep up with and extract useful insights from many news articles. Although many preceding works focused on the sentiment extracted in the news, the results were intended for everyone. The sentiments based on a user's queries are needed to provide customized service. Hence, this paper proposed a system integrated into a chatbot to automatically understand users' queries and recommend sentiments based on news articles. The main objective is to provide entrepreneurs, especially those considering international trade and investment, with the sentiments embodied in the latest news articles to help them keep up with the business and economic trends relevant to them. The methodology is based on deep learning and transfer learning. A pre-trained deep learning model was fine-tuned for natural language processing tasks to perform sentiment analysis in news articles. A survey questionnaire was used to measure the effectiveness of the system. The survey result showed that most users agreed with the predicted sentiments from the system.**

***Keywords*—Sentiment analysis; deep learning; pre-trained model; natural language processing**

I. Introduction

One of the important sources of business information for entrepreneurs is news articles. News, especially business and economics news, has become one of the factors that entrepreneurs consider when deciding to trade or expand the business to other countries. The news could be more than just a medium to report on what has happened or carry information about the topics or entities discussed in the articles. They could also carry sentiments or trends toward the main content, such as products or companies mentioned in the articles. In other words, news conveys how the authors and people feel about the entity. This public sentiment is crucial to business because it tells how the public feels about the product or service. It also significantly influences the future of the product or service industries. But with the advancement in internet technology, the amount of news articles is increasing rapidly. Therefore, it has become a challenge for entrepreneurs to keep up with the rapidly increasing number of news sources, filter out unreliable sources, find all relevant information they need, and extract sentiments manually.

Sentiment analysis is a natural language processing (NLP) technique used to determine and extract subjective information in a text. The information could be people's opinions, sentiments, or attitudes towards entities. Due to the rapid growth of data on social media, sentiment analysis has become one of the most active research topics in NLP. Moreover, due to its superior performance in many application domains, deep learning has gained popularity, thanks to advanced cloud-based technology and increasing computing power. Among the success of deep learning in various application domains, deep learning has been used in sentiment analysis [1-3]. Many deep learning models can be applied or adapted to NLP datasets to attain high accuracy [4]. Therefore, this paper focuses on applying deep learning to automatically analyze many news articles and recommend sentiments based on news articles. The suggested sentiment can be considered an indicator that reflects consumers' attitudes and economic outlook toward the specific product in the country. Therefore, it can help a user make proper business decisions.

There are two main contributions to this paper. First, a deep-learning-based model is proposed to automatically understand users' queries and analyze and recommend sentiments based on current news articles. Second, the proposed model was integrated into a chatbot to provide an end-to-end solution in a business and economic domain so that the model could be tested in practice. As a result of this system, users can stay up to date on business and economic trends relevant to their specific fields.

The rest of the paper is organized as follows. First, an overview of the related works is presented in Section II. Next, the methodology of the proposed system is discussed in Section III. Then Section IV shows the results and findings. Finally, Section V gives conclusions regarding this work.

II. Related Work

Sentiment analysis has begun to be used in economics and finance recently. Many studies are concentrating on using sentiment analysis for stock prediction. Several works [5-7] analyzed tweets' moods or opinions and used machine learning methods for prediction. [5] studied whether the collective mood states extracted from the Twitter feed can be used to predict the value of the Dow Jones Industrial Average (DJIA). They used a Self-Organizing Fuzzy Neural Network to predict the changes in DJIA closing values. The results showed that the prediction accuracy was significantly improved when including a calm sentiment. [6] calculated a sentiment score of each tweet using a dictionary-based method and created feature vectors of sentiment scores to train a support vector machine (SVM) to classify the stock trend. The best accuracy was 90.34%. [7] also used a dictionary-based method with eight sentiments to analyze Twitter data. However, the results showed that adding sentiment data did not significantly improve accuracy.

Recent works [8-10] have developed a prediction model based on a deep learning approach. [8] used deep learning models to extract features from news headlines and predict stock prices. A Convolutional Neural Network (CNN) model was used to transform the sequence of words in the news title to the level of sentiment. The sentiment and other technical indicators were inputs to the Long Short-term Memory (LSTM) model to predict the price movement. The model achieved a 97.66% accuracy rate. [9] employed a similar approach where they extracted sentiments in the news content on Sina Weibo, China's largest online social network. They then input the sentiment features and technical indicators into an RNN-boost model to predict the stock volatility in the Chinese stock market. [10] enhanced LSTM by utilizing an attention mechanism to make predictions on the final output and informative outputs from hidden states. The results show that the attention-based LSTM improved prediction accuracy and reduced computational time.

Several other papers have focused on the economic sentiment embodied in the news and social media, especially Twitter. The author in [11] proposed a new technique to measure economic sentiment embodied in the news. Unlike survey-based economic sentiment measures, their index relies on extracting sentiment from news articles. The technique was applied to two applications. In the first application, they use their news sentiment index to predict consumer sentiment based on surveys. They found that the news sentiment is strongly predictive of Michigan Consumer Sentiment Index and the Conference Board's Consumer Confidence Index. Second, they investigated how the macro-economic response to sentiment shocks. They found that positive sentiment shocks increase consumption, output, and interest rates and temporarily reduce inflation. The author in [12] build a model to predict the sentiment index of a company based on news about that company. The authors evaluated their model by measuring the model's accuracy to predict the company's stock price movement. The result showed that the model has an average accuracy score of around 70.1%. The author in [13] proposed an automatic news chatbot that provides a variety of news articles organized into chatrooms based on news topics. When a user enters a chatroom, the chatbot provides the latest news articles on a given topic. A user can also ask specific questions regarding the topic, which a chatbot will answer with short sentences and provide a link to an article containing the answer. The drawback is a user cannot start a conversation by asking a specific question. Instead, a user must pick one of the topics to have conversations about that topic.

While most of the preceding work focused on the sentiment extracted in the news content where the results were intended for everyone, such as the consumer sentiment or sentiments for stock predictions, the sentiments based on a user's queries are needed to provide a customized service to the users. As a result, this work proposed a system that automatically understands users' queries and recommends sentiments based on current news articles. The system was integrated into a chatbot to provide easy access to the system for users. The system would help users to keep up with the latest business and economic trends in their fields of interest.

III. Methodology

This section describes the system architecture and how it processes users' queries to suggest product sentiments to users. The system consists of four components: a conversational interface, a keyword extraction module, a news search module, and a sentiment analysis module, as shown in Fig. 1. First, a user sends a query through a conversational interface. Next, the query is sent through the system. After receiving the query, the system extracts a keyword from the query and gathers news articles from trusted sources related to the keywords. It then analyzes the articles and suggests the sentiment. The news articles' list and their sentiment are then sent back to a user. The details of each component are explained in the following subsections.

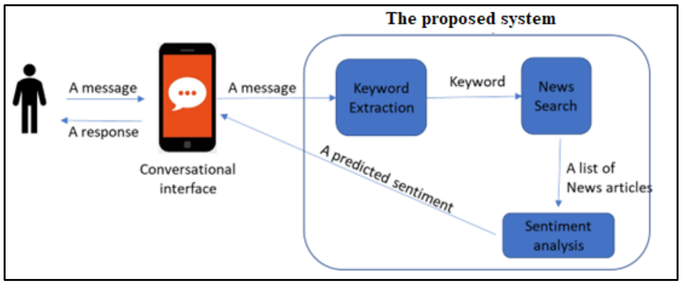
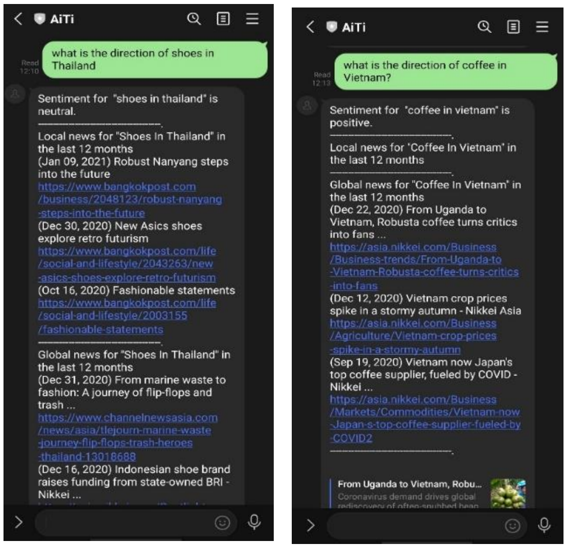


Fig. 1. The Architecture of the Proposed System.

A. The Conversational Interface

The first component is a conversational interface acted as a chatbot to interact with users. The LINE messenger app was used as the conversational interface for this work because it is Thailand's most used messaging app [14]. Thus, the system can easily be accessible to Thai users.



what is the direction of shoes in Thailand

Sentiment for "shoes in thailand" is neutral.
------------------------------------.
Local news for "Shoes In Thailand" in the last 12 months
(Jan 09, 2021) Robust Nanyang steps into the future
https://www.bangkokpost.com/business/2048123/robust-nanyang-steps-into-the-future
(Dec 30, 2020) New Asics shoes explore retro futurism
https://www.bangkokpost.com/life/social-and-lifestyle/2043263/new-asics-shoes-explore-retro-futurism
(Oct 16, 2020) Fashionable statements
https://www.bangkokpost.com/life/social-and-lifestyle/2003155/fashionable-statements
------------------------------------.
Global news for "Shoes In Thailand" in the last 12 months
(Dec 31, 2020) From marine waste to fashion: A journey of flip-flops and trash ...
https://www.channelnewsasia.com/news/asia/tlejourn-marine-waste-journey-flip-flops-trash-heroes-thailand-13018688
(Dec 16, 2020) Indonesian shoe brand raises funding from state-owned BRI - Nikkei ...

(a)

what is the direction of coffee in Vietnam?

Sentiment for "coffee in vietnam" is positive.
------------------------------------.
Local news for "Coffee In Vietnam" in the last 12 months
------------------------------------.
Global news for "Coffee In Vietnam" in the last 12 months
(Dec 22, 2020) From Uganda to Vietnam, Robusta coffee turns critics into fans ...
https://asia.nikkei.com/Business/Business-trends/From-Uganda-to-Vietnam-Robusta-coffee-turns-critics-into-fans
(Dec 12, 2020) Vietnam crop prices spike in a stormy autumn - Nikkei Asia
https://asia.nikkei.com/Business/Agriculture/Vietnam-crop-prices-spike-in-a-stormy-autumn
(Sep 19, 2020) Vietnam now Japan's top coffee supplier, fueled by COVID - Nikkei ...
https://asia.nikkei.com/Business/Markets/Commodities/Vietnam-now-Japan-s-top-coffee-supplier-fueled-by-COVID2
------------------------------------.

From Uganda to Vietnam, Robu...
Coronavirus demand drives global rediscovery of often-snubbed bean

(b)

Fig. 2. The Conversational Interface. (a): Regular Response (b): Response with Empty List of Local News.

The conversational interfaces are shown in Fig. 2(a) and 2(b). Fig. 2(a) shows a response to the LINE messenger app when a user sends a message, "What is the direction of shoes in Thailand?". The first line in the response message is the sentiment of the keywords "shoes" and "Thailand." After the first line, it shows a list of local and global news articles related to the keywords. The local news is the news from local news providers such as bangkokpost.com or vietnamnews.vn. The global news is the news from global news providers such as channelnewsasia.com or asia.nikkei.com. If there is no related news, the list will be empty, as shown in Fig. 2(b) where the list of local news is empty.

The system suggests three types of sentiment: positive, neutral, and negative. The interpretation of the sentiments of keywords is as follows. The positive and negative sentiments mean the keywords have a positive or negative economic outlook. While the neutral sentiment means the keywords have neither a positive nor negative economic outlook.

B. The Keyword Extraction Module

The main objective of this module is to extract keywords in a user's query and pass the keywords to the news search module for searching relevant news articles. The keywords can be a single word or a combination of words in a query. Fig. 3 displays an overview of the input and output of the keyword extraction module. First, a user sends a message, "What is the trend of tourism in Myanmar?" via a conversational interface. Next, the message is sent to the keyword extraction module. The module then extracts keywords; in this example, the extracted keywords are "tourism" and "Myanmar." The News search module then uses this keyword to search for relevant news articles. If the module cannot identify any keywords, the system will tell a user that no keywords can be identified and ask a user to try a different query.

The module first tokenizes the sentences and identifies parts of speech (POS) tags to extract keywords. The keyword in the message is a combination of words tagged as nouns and proper nouns, both singular and plural. The result of tokenizing and POS tagging the sentence “What is the trend of tourism in Myanmar?' looks like this:

('What', 'WP')

('is', 'VBZ')

('the', 'DT')

('trend', 'NN')

('of', 'IN')

('tourism', 'NN')

('in', 'IN')

('Myanmar', 'NNP')

('?', '.')

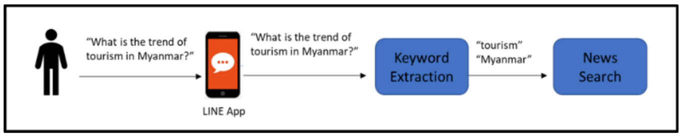

"What is the trend of tourism in Myanmar?"
LINE App
"What is the trend of tourism in Myanmar?"
Keyword Extraction
"tourism" "Myanmar"
News Search

Fig. 3. An Overview of the Input and Output of the Keyword Extraction Module.

The tagging result has three nouns: 'trend,' 'tourism,' and 'Myanmar' in the sentence. After filtering out the word 'trend,' the keyword is "tourism Myanmar."

This work used the NLTK toolkit [15] for tokenizations and POS tagging. NLTK is a leading tool for working with NLP in Python. It provides various libraries for text processing, including tokenizations and POS tagging. Tokenization is how a sentence is broken into words and punctuations. For example, "What is the trend of tourism in Myanmar?" is tokenized as ['What', 'is', 'the', 'trend', 'of', 'tourism', 'in', 'Myanmar', '?']. After that, the module classifies words into parts of speech and labels them accordingly using a POS tagger. The POS tagger uses a UPenn Tagset, as shown in Table I.

TABLE I. A LIST OF PART-OF-SPEECH TAGS

Tag	**Description**
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential *there*
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
M.D.	Modal
N.N.	Noun, singular, or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun
R.B.	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	*to*
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund, or present participle
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WP$	Possessive wh-pronoun
WRB	Wh-adverb

C. *News Search Module*

This module searches for relevant articles across trusted news sources using the extracted keyword from the keyword extraction module. Two lists of trustworthy news sources were curated: a local and global list. The local list contains high-reputable local news sources from each country, while the global list contains high-reputable news sources covering news mainly from Asia and worldwide. The local news sources are as follows:

- https://en.vietnamplus.vn
- https://english.vov.vn
- https://aecnewstoday.com
- www.irrawaddy.com
- https://elevenmyanmar.com
- https://english.cambodiadaily.com
- www.phnompenhpost.com
- https://laotiantimes.com
- https://www.bangkokpost.com/

The global news sources are as follows:

- asiatimes.com
- www3.nhk.or.jp
- https://asia.nikkei.com
- https://www.asiaone.com
- https://www.businessnewsasia.com
- https://www.channelnewsasia.com
- http://annx.asianews.network
- www.businesstimes.com.sg

Google Programmable Search Engine (formerly known as Google Custom Search) was used to search across a collection of local and global news sources. A custom search JSON API was created with the Google Programmable Search Engine to retrieve search results automatically in JSON format via RESTful requests. In addition, this module was set to return news articles published within 12 months.

D. *Sentiment Classification Module*

This module classifies a sentiment related to the user's keyword. The sentiment of the keyword is based on sentiments of its related news articles. The keyword's sentiment depends on how the news providers and people feel about the content the keyword was discussed. This module consists of two models: a deep learning model and a classifier, as shown in Fig. 4.

Fig. 4 shows how the Sentiment Classification Module works. This example uses a user query "What is the trend of tourism in Myanmar?" where the news search module returns a list of news articles as follows:

- Article 1: Week in Review: Tourist arrivals fall 75% in 2020.
- Article 2: Myanmar sees 75% drop in tourist arrivals.
- Article 3: Mandalay eyes tourism revival post-pandemic.
- Article 4: Mandalay prepares to reopen hotels, revive tourism.

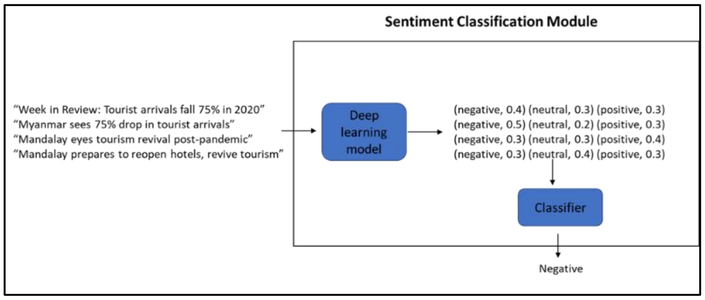

Sentiment Classification Module
"Week in Review: Tourist arrivals fall 75% in 2020"
"Myanmar sees 75% drop in tourist arrivals"
"Mandalay eyes tourism revival post-pandemic"
"Mandalay prepares to reopen hotels, revive tourism"
Deep learning model
(negative, 0.4) (neutral, 0.3) (positive, 0.3)
(negative, 0.5) (neutral, 0.2) (positive, 0.3)
(negative, 0.3) (neutral, 0.3) (positive, 0.4)
(negative, 0.3) (neutral, 0.4) (positive, 0.3)
Classifier
Negative

Fig. 4. An Example of How the Sentiment Classification Module Works.

After receiving a list of news articles, the sentiment classification module performs two steps. First, it calls a deep learning model to calculate sentiment scores for each article. There are three sentiments: negative, neutral, and positive. Hence each article is assigned three scores for the three sentiments. The result from the first step would look like this:

- Article 1: (negative, 0.4) (neutral, 0.3) (positive, 0.3).
- Article 2: (negative, 0.5) (neutral, 0.2) (positive, 0.3).
- Article 3: (negative, 0.3) (neutral, 0.3) (positive, 0.4).
- Article 4: (negative, 0.3) (neutral, 0.4) (positive, 0.3).

Second, it classifies a sentiment for the keywords by choosing the sentiment with the highest average scores from all news articles. From the example, the averages of the three sentiments from four articles are below:

- (negative, 0.375) (neutral, 0.3) (positive, 0.325)

The averages show that the negative sentiment has the highest score, so it classifies the sentiment's keywords as "Negative". The system then sends the result back to the user, saying, " The tourism trend in Myanmar is Negative.".

1) Sentiment classification module: Many works provide methods and tools for sentiment analysis. Recently, most meth-ods have been based on deep learning with transfer learning. Deep learning is a sub-field of machine learning methods based on neural networks. The networks typically have three or more layers, performing feature extraction and transformation via a cascade of multiple layers of linear and nonlinear processing nodes. The lower hidden layers near the input layer learn simple features, while the higher layers learn complex features derived from lower layers.

A transfer learning is a machine learning technique where a model is trained for one task and applied to a different but related task. The common approach to using transfer learning for deep learning is to use a pre-trained model. A pre-trained model is chosen from available models typically released by research institutes for this approach. The pre-trained model is then trained and fine-tuned on a new dataset for the task of interest.

Many pre-trained deep learning models can be employed for the task. Hence, four models were tested in the experiments, Bert, RoBERTa, XLM-R, and XLNet. The best model would be selected for the system. Google's Bert [16] is a bi-directional model pre-trained on an unlabeled text that can be used to create a wide range of tasks by fine-tuning with just one additional output layer. RoBERTa [17] extends the original BERT model where the researchers fine-tune the original BERT model with a huge dataset and improved input representation. XLM-R [18] is a cross-lingual language model trained with MLM (Masked Language Modeling) on one hundred languages and terabytes of texts. Finally, XLNet [19] is an auto-regressive language model which uses the context word to predict the next word. The model has also addressed some drawbacks of BERT and outperformed BERT in many tasks.

A linear layer was added on top of the pooled output to calculate sentiment scores, as shown in Fig. 5. Next, all models were trained and fine-tuned for a classification task where the outputs are probabilities of three classes: positive, negative, and neutral.

2) The dataset: The dataset used to train and fine-tune models was curated by [20]. The dataset contains about 5000 news titles from financial news resources, where each title was labeled as positive, negative, or neutral by annotators with good backgrounds in business and investment. The classes reflect an investor's perception of business and market conditions on the news. The positive class means the news title appears to positively influence the market and vice versa for the negative class. If the news title reflects neither positive nor negative influence, the news title is considered neutral. Each news title was trimmed to have the maximum sequence length of 256 for fast processing in a real-time environment. Examples of news titles and their sentiments are shown in Table II.

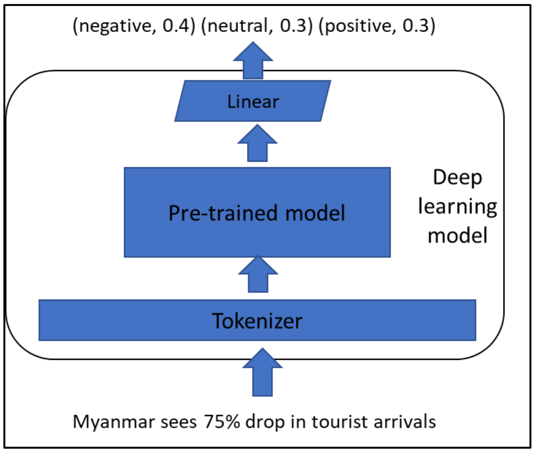

(negative, 0.4) (neutral, 0.3) (positive, 0.3)
Linear
Pre-trained model
Deep learning model
Tokenizer
Myanmar sees 75% drop in tourist arrivals

Fig. 5. The Architecture of the Proposed Deep Learning Model.

TABLE II. EXAMPLES OF NEWS TITLES AND THEIR SENTIMENTS

News Title	Sentiment
According to Gran, the company has no plans to move all production to Russia, although that is where the company is growing.	neutral
Technopolis plans to develop in stages an area of no less than 100,000 square meters to host companies working in computer technologies and telecommunications, the statement said.	neutral
With the new production plant, the company would increase its capacity to meet the expected increase in demand and would improve the use of raw materials, and therefore increase the production profitability.	positive
According to the company's updated strategy for the years 2009-2012, Basware targets a long-term net sales growth in the range of 20 % -40 % with an operating profit margin of 10 % -20 % of net sales.	positive
The international electronics industry company Elcoteq has laid off tens of employees from its Tallinn facility; contrary to earlier layoffs the company contracted the ranks of its office workers, the daily Postimees reported	negative
A TinyURL link takes users to a scamming site promising that users can earn thousands of dollars by becoming a Google (NASDAQ: GOOG) Cash advertiser.	negative

IV. RESULTS AND DISCUSSION

This section shows the performance of different pre-trained deep learning models for model selection and the system's performance based on a user survey. Error analysis and some discussions are also provided.

A. Model Selection

Four deep learning models were tested for model selection. The models were Bert, RoBERTa, XLM-R, and XLNet. The Transformers and FastAI libraries [21, 22] with the default architecture were used to implement the deep learning models where weights were initiated randomly. First, the dataset was divided into a train and test set with a split ratio of 60:40. The training set was used to fine-tune the pre-trained models, and the test set was used to select the best model. Next, the models were trained with the Adam optimization algorithm with an accuracy rate as a metric. Adam optimization algorithm is an algorithm to update the weights in neural networks during training. The Adam optimization algorithm was chosen because it works well in many empirical results and is recommended as the default algorithm [23, 24]. The key metrics to evaluate the performance of the models are accuracy rate and computation time on the test set.

From Table III, the accuracy rates of Bert, RoBERTa, XLM-R, and XLNet are 0.8500, 0.8633, 0.7367, and 0.8667, respectively. From the results, RoBERTa is tied with XLNet regarding the accuracy rate, but it took the least computational time, so RoBERTa was chosen for the system.

TABLE III. EXPERIMENTAL RESULTS ON A MODEL SELECTION

Model	Accuracy Rate	Computational Time (Seconds)
Bert	0.8500	0.0428
RoBERTa	0.8633	0.0378
XLM-R	0.7367	0.3742
XLNet	0.8667	0.0399

B. Performance Evaluation

There is no public ground truth for evaluating the accuracy of the predicted sentiments of keywords because a sentiment of a keyword depends on the current economics, and it can change as time goes by. For this reason, a user survey was employed to measure the system's effectiveness from users' perspectives. First, the participants were asked to make a few queries using keywords of their choices and give feedback on the results. After that, users were asked two questions:

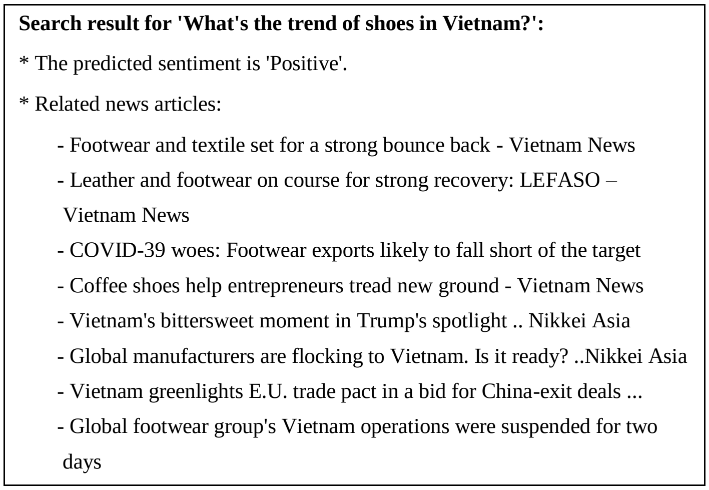
1) On a scale of 1 – 5, with 1 being highly irrelevant and 5 being highly relevant, how relevant each news article is to your keyword?

2) What sentiment would you assign to your keyword from your perspective?

The first question is to measure the relevancy of the news articles to the keyword. Participants were asked to rate an individual news article. A relevance rating for the keyword was calculated by averaging the ratings of all of its news articles. The second question measures the accuracy of the keyword sentiment from the users' perspective. It is used as a keyword accuracy rate, the fraction of predictions the system got right according to users' perspectives.

For example, a user searching for a 'What's the trend of shoes in Vietnam?' result is shown in Fig. 6. In the figure, the predicted sentiment for the keyword is "Positive," where the extracted keyword for the user's query is 'shoes Vietnam'. The system also returned related recent news articles to the user. Each title contains a link to its original news source.

Next, the participants were asked to rate each article in terms of relevancy to the keywords. The rating scale is 1 to 5. Table IV shows a news relevance rating for each news article and the news relevance score, which is an average of all ratings. For this example, the news relevance score is 3.88, indicating that the news articles are somewhat relevant to the keyword. The participants were also asked what sentiment they would assign to the keyword based on the news and their perspective. For example, for the keyword 'shoes in Vietnam', the user assigned a 'Positive' sentiment the same as the sentiment predicted by the system.



Search result for 'What's the trend of shoes in Vietnam?':

* The predicted sentiment is 'Positive'.

* Related news articles:

- Footwear and textile set for a strong bounce back - Vietnam News
- Leather and footwear on course for strong recovery: LEFASO – Vietnam News
- COVID-39 woes: Footwear exports likely to fall short of the target
- Coffee shoes help entrepreneurs tread new ground - Vietnam News
- Vietnam's bittersweet moment in Trump's spotlight .. Nikkei Asia
- Global manufacturers are flocking to Vietnam. Is it ready? ..Nikkei Asia
- Vietnam greenlights E.U. trade pact in a bid for China-exit deals ...
- Global footwear group's Vietnam operations were suspended for two days

Fig. 6. Search Result for a user's Keyword.

TABLE IV. RELEVANCE RATINGS

Keyword	News Title	News Relevance Rating
shoes in Vietnam	Footwear and textile set for a strong bounce back - Vietnam News	5
	Leather and footwear on course for strong recovery: LEFASO - Vietnam News	5
	COVID-39 woes: Footwear exports likely to fall short of the target	3
	Coffee shoes help entrepreneurs tread new ground - Vietnam News	5
	Vietnam's bittersweet moment in Trump's spotlight - Nikkei Asia	3
	Global manufacturers are flocking to Vietnam. Is it ready? - Nikkei Asia	3
	Vietnam greenlights E.U. trade pact in bid for China-exit deals ...	2
	Global footwear group's Vietnam operations were suspended for two days	5
	Relevance rating	**3.88**

Fifteen users participated in the survey. Seven of the users were entrepreneurs, and the remaining users were senior students with a major in international business management. Each user made 3 to 4 queries. The total number of queries was 40. The survey results show that the average relevance rating is 2.66 out of 5, and the accuracy rate of the keyword sentiment is 0.35. The average relevance rating and accuracy rate for each sentiment category were also computed, as shown in Table V. The results look promising, with a 62.50% accuracy rate for the positive articles and a 2.93 relevance rating for the negative articles.

TABLE V. THE RELEVANCE RATING AND THE ACCURACY RATE BY CATEGORY

Sentiment	Relevance Rating	Accuracy Rate
Negative	2.9342	0.3636
Neutral	2.5133	0.2380
Positive	2.7142	0.6250

C. Error Analysis

After the evaluation process, classification errors were examined to identify how to improve the system. First, a confusion matrix was constructed to see what categories the system misclassified and what categories had been predicted correctly. From the confusion matrix in Table VI, the system predicted Neural sentiments the most (21/40), whereas users assigned Positive sentiments the most (24/40). So the results of 14 keywords that the system predicted Neutral sentiments, but users assigned Positive sentiments were examined. The examination result shows that a few articles sound positive but were assigned Neutral sentiment by the system. The Neutral sentiment also received the lowest relevance rating and accuracy rate, as shown in Table V. These results may be due to the class imbalance in the dataset used to train the model, where 59% of articles were Neutral. In comparison, only 12% and 28% of articles were Negative and Positive, respectively. Based on this finding, the system's performance in the Neutral category could be improved by employing a resampling technique to deal with an unbalanced dataset.

TABLE VI. CONFUSION MATRIX

		User Sentiment			**Total**
		Negative	**Neutral**	**Positive**	
Predicted Sentiment	**Negative**	4	2	5	11
	Neutral	2	5	14	21
	Positive	2	1	5	8
Total		8	8	24	40

Second, the queries that received the correct sentiments were analyzed. The result shows that they also received a higher news relevance rating. The finding may indicate that higher relevant news articles will likely improve the sentiment accuracy. To test if higher relevant news articles would help boost the accuracy rate, queries whose average relevance rating is less than 3 were removed and re-calculated the accuracy rate. There are 22 queries whose average relevance rating is less than 2. After removing those queries, the accuracy rate went up to 0.5. From the analysis, the search engine should be improved to return more relevant news articles so that the algorithm has better information to calculate a sentiment.

Third, since the accuracy of keyword sentiments depends on the accuracy of predicted sentiments of news articles related to the keyword, the senior students who participated in the survey were asked to rate sentiments of the news articles in the search results from the survey. There were 268 news articles from 40 queries. Next, the accuracy rate and a macro-F1 score of the news sentiment were computed by comparing the predicted sentiment classifications from the model to the sentiments rated by the users. The accuracy rate and a macro-F1 score are 0.6679 and 0.4343, respectively. These results show that the system could accurately classify news sentiments, but those articles were irrelevant to their keywords. These results also emphasize the need to improve the system's search engine.

D. Discussion

The accuracy rate and a macro-F1 score of the news sentiment from the system are 0.6679 and 0.4343, respectively. These results are similar to the result in [11], where the authors proposed a new technique to measure economic sentiment embodied in the news. In [11], the macro-F1 score was calculated using the 100-article test set for which they compared predicted sentiments from various models to human-provided sentiments. The lowest macro-F1 score was 0.406 from using Lexicon only, and the highest macro-F1 score was 0.525 from their proposed method. According to these macro-F1 scores, sentiment analysis is still very challenging. The majority of the current sentiment analysis techniques are data-driven machine learning techniques. Hence they have limitations in terms of data size and inconsistency of ground truth [25]. Furthermore, labeling needs to be done by humans, and human ability to label large volumes of data is limited. Moreover, there is a great deal of subjective opinions when it comes to business news. A bad situation for one business might be good for another business. As a result, it is essential to provide clear and concise instructions to produce high-quality annotations. In this study, the participants were asked to make queries related to their businesses or expertise to evaluate results from those businesses' perspectives.

V. CONCLUSION

This paper presents a methodology to automatically understand a user's query about a product and provide its sentiment embodied in news articles. The system is based on deep learning and transfer learning to build a model using a pre-trained deep learning model fine-tuned for sentiment analysis in news articles. The news articles are automatically searched and collected by the news search module from the lists of trustworthy news sources both locally and globally to ensure the quality of news articles. Finally, the model was integrated into a chatbot and tested in practice. The satisfaction survey shows participants agreed with the results, with a relevance rating of 2.66 and an accuracy rate of 35%. The evaluation by category shows that the positive articles received the highest accuracy rate of 62.50%, while the negative articles received the highest relevance rating of 2.93. In the future, the news search module could be improved to return relevant search results with higher precision by adding more capabilities such as semantic understanding to understand user intention better. Moreover, the sentiment classification module could be improved by re-training the deep learning model with different parameters.

REFERENCES

[1] Chen, Y., Convolutional neural network for sentence classification. 2015, University of Waterloo.

[2] Dos Santos, C. 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.

[3] Guggilla, C., T. Miller, and I. Gurevych. CNN-and LSTM-based claim classification in online user comments. In Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers. 2016.

[4] Zhang, L., S. Wang, and B. Liu, Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018. 8(4): p. e1253.

[5] Bollen, J., H. Mao, and X. Zeng, Twitter mood predicts the stock market. Journal of computational science, 2011. 2(1): p. 1-8.

[6] Meesad, P. and J. Li. Stock trend prediction relying on text mining and sentiment analysis with tweets. In 2014 4th World Congress on Information and Communication Technologies (WICT 2014). 2014. IEEE.

[7] Porshnev, A., 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.

[8] Liu, Y., et al. Stock price movement prediction from financial news with deep learning and knowledge graph embedding. In Pacific rim knowledge acquisition workshop. 2018. Springer.

[9] Chen, W., et al., Leveraging social media news to predict stock index movement using RNN-boost. Data & Knowledge Engineering, 2018. 118: p. 14-24.

[10] Jin, Z., Y. Yang, and Y. Liu, Stock closing price prediction based on sentiment analysis and LSTM. Neural Computing and Applications, 2020. 32(13): p. 9713-9729.

[11] Shapiro, A.H., M. Sudhof, and D.J. Wilson, Measuring news sentiment. Journal of econometrics, 2020.

[12] Chowdhury, S.G., S. Routh, and S. Chakrabarti, News analytics and sentiment analysis to predict stock price trends. International Journal of

Computer Science and Information Technologies, 2014. 5(3): p. 3595-3604.

[13] Laban, P., J. Canny, and M.A. Hearst, What's The Latest? A Question-driven News Chatbot. arXiv preprint arXiv:2105.05392, 2021.

[14] Corporation, N., 2020 NAVER Annual Report. 2020: Online.

[15] Loper, E. and S. Bird, Nltk: The natural language toolkit. arXiv preprint cs/0205028, 2002.

[16] 1Devlin, J., et al., Bert: Pretraining of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[17] Liu, Y., et al., Roberta: A robustly optimized Bert pretraining approach. arXiv preprint arXiv:1907.11692, 2019.

[18] Conneau, A., et al., Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116, 2019.

[19] Yang, Z., et al., generalized autoregressive pretraining for language understanding; 2019. Preprint at https://arxiv.org/abs/1906.08237 Accessed June, 2021. 21.

[20] Malo, P., et al., Good debt or bad debt: Detecting semantic orientations in economic texts. Journal of the Association for Information Science and Technology, 2014. 65(4): p. 782-796.

[21] Howard, J. and S. Gugger, Fastai: a layered API for deep learning. Information, 2020. 11(2): p. 108.

[22] Wolf, T., et al. Transformers: State-of-the-Art Natural Language Processing. 2020. Online: Association for Computational Linguistics.

[23] Kingma, D.P. and J. Ba, Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[24] Ruder, S., An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747, 2016.

[25] Wankhade, M., Rao, A.C.S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev (2022). https://doi.org/10.1007/s10462-022-10144-1.

Sentiment Analysis using Term based Method for Customers' Reviews in Amazon Product

Thilageswari a/p Sinnasamy, Nilam Nur Amir Sjaif
Razak Faculty of Technology and Informatics
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia

***Abstract*—Customers' review in Amazon platform plays an important role for making online purchase decision making, however the reviews are snowballing in E-commerce day by day. The active sharing of customers' experience and feedback helps to predict the products and retailers' quality by using natural language processing. This paper will focus on experimental discussion on Amazon products reviews analysis coupled with sentiment analysis using term-based method and N-gram to achieve best findings. The investigation of sentiment analysis on amazon product gain more valuable information on related text to solve problem related services, products information and quality. The analysis begins with data pre-processing of Amazon products reviews then feature extraction with POS tagging and term-based concept. e-Commerce customer's reviews normally classify different experience into positive, negative and neutral to judge human behavior and emotion towards the purchase products. The major findings discussed in this journal will be using four different classifier and N-grams methods by computing accuracy, precision, recall and F1-Score. TF-IDF method with N-gram shows unigram with Support Vector Machine learning with highest accuracy results for Amazon product customers' reviews. The score reveals that Support Vector Machine for unigram achieved 82.27% for accuracy, 82% precision, 80% Re-call and 72% F1-Score.**

Keywords*—*Sentiment analysis; e-commerce; term based; n-gram

I. Introduction

The huge growth of opinion sharing platform in E-commerce such as Amazon, E-bay, Shopee, Zalora and Lazada could facilitates the customers to understand better about products that been available in online platform [1] [2]. Sharing customer's experience and feedback help to predict the products and retailer quality by using natural language processing. Furthermore, customer's experience rating (1 to 5 stars) and reviews also play important role to express customers satisfaction namely 4 and 5 score represent positive attitude, score 3 represent neutral and score 1 and 2 represent negative attitude [3] [4]. Product star ratings in E-commerce used to evaluate product quality information and online customer behavior from various dimensions whether it could be positive or negative.

Amazon is one popular online platform for buying and selling online products purchase by customers [2] [5]. Customers' review and rating in Amazon platform plays important role to make a purchase decision however the reviews have grown remarkably in e-Commerce. Naturally, it consumes plenty of time to digest and explore huge amount of reviews to make a right decision [2] [3] [6] [7]. Other than that, customer may also get indecisive to make right purchase decision especially the 3-stars in Amazon represent neutral. In addition, there are also inadequate summary of dominant reviews from E-commerce platforms for satisfied customers [8]. The semantic words in reviews also create noise to the result and low frequency problem.

The aim of this paper is to explore on quality of text representation and categorization with better classifiers as to measure and judge the important raw information. Therefore, sentiment analysis help predicts the polarity sentiments in reviews which could be positive, neutral or negative. Sentiment analysis explores expression and emotion of customers which feeds extract important information from customers' reviews [6]. Sentiment classification strategy evaluate and described the text in reviews with aspect level by identifying interest words from reviews using text mining method [6] [9], whereby sentiment analysis approach machine learning and dictionary based method [10]. Machine learning method classifies the text and dictionary method identifies polarity of words. In this case, text mining would drive the information to organize text as positive, negative and neutral. Text mining retrieves the relevant information to segregate between structured and unstructured data [11] [12]. This paper will present experimental result and analysis using term based method using data from Amazon products. The objective of this paper is to reveal the insightful meaning and identify the polarity pattern from unstructured data using sentiment analysis. It includes method as data pre-processing, feature extraction, sentiment polarity prediction and classification.

This paper would be sorted out such a manner that after this introduction, Section II comprises of related works it covers some of previous studies on sentiment analysis. Later in Section III, it elaborates text mining methods. The following Section IV will explain in detail of the approach on data and method as solution to this paper summary. The methodology used in this experiment as describe deeply in this section. Next, Section V present summary of result on discussion obtained from experiment. Section VI discuss on contribution of this paper to customers using e-Commerce. Finally, the last section concludes the findings of paper and acknowledgment.

II. Related Work

This section presents several studies with different approaches to analyze sentiments in review. Many researchers have experience to work with customers' reviewers in using sentiment analysis. Naturally Customers' reviews are very important to identify product satisfaction, product features and services for convenient customers [3] [4] [13]. The sentiment analysis result from various method helps provide effective purchase decision and reducing search cost in E-Commerce. Sentiment analysis polarize the review into positive, neutral and negative with natural language processing for judging human behavior and attitude [6]. Authors in [14] have worked with sentiment analysis using Support Vector Machine (SVM) to analyze reviews in tweets. The survey results present several techniques and methods of sentiment analysis which provides different results but if combine both methods then it produces better result. The main factors of best result are selection of methodology for data preprocessing, feature extraction phase and ratio between training and testing. Hence, author Minu with other researchers [15] has conduct experiments on mobile products reviews using machine learning algorithms and exploring the result with different datasets. POS tagging used to recognize the reviews noun, adjective, verb and adverb. The aspects are than polarize into positive and negative using TextBlob and machine learning approaches such as K-Nearest Neighbour, Support Vector Machine, Multinomial Naïve Bayesian and Bernoulli Naïve Bayesian used for predict the result. The best accuracy of mobile reviews is given by K-Nearest Neighbour and Bernoulli Naïve Bayesian. Furthermore, the sentiment analysis also implemented on different language for analyze various machine learning algorithm with N-gram as feature extraction [16]. Arabic opinion from twitter collected and associate data pre-processing with the phrases for noise reduction. N-gram which is number of terms used to search word such unigram (1), bigram (2) and trigram (3) used as feature extraction to generate frequent appealing words. By applying machine learning approaches the result present that PA and RR using unigram, bigram and trigram present best result as 99.96%. On other hand, online customer reviews from Tokopedia been analyze for understanding service quality level using sentiment analysis [17]. Whereby the unstructured data performed starting with data pre-processing, TF-IDF implemented for identifying frequent itemset and reduce noisy from dataset. Finally, Naïve Bayes classifiers used for measure accuracy, recall and precision. The result shows Naïve Bayes present very good result for classifying sentiment and the overall methods used are more effective to perform better. More details on text mining methods been discussed in Section III.

III. Text Mining Methods

Some prominent researchers have used text mining methods in E-commerce such as the Amazon reviews from large textual database in the form of structured and unstructured data [12] [18]. Text mining includes process such as data collection, data pre-processing, feature extraction, classification and measurement. Even Machine learning tools been utilized to explore text mining method to organize and categorize data in further details. The process data used for predictive analysis, business application, business intelligence, decision support system and data warehouse where there is a need to refer customers' requirement [19]. There are four method of text mining are discussed in this section such as Term based, Phrase based, Concept based and Pattern based.

A. Term Based

The dataset filtered and stemmed to obtain frequency of term used to represent as document is presented with term-by-frequency matrix [18]. While Term base method contributes terms into documents as to discover weights for each describe terms. Term frequency and inverse document frequency (TF-IDF) model generally been used to calculate frequent word in a document with inverse proportion of word whereby the model converts textual data to Vector Space Model (VSM) [20].

TF : Term Frequency refer to number of term in a documents [18] [21]. TF calculated as [22] :

$$\mathrm{TF}(t,d) = \frac{n_t}{N_{(T,d)}} \qquad (1)$$

Nt represent number of frequency of term t in a document d, *N(T,d)* shows sum of total terms *T* in document.

IDF : Inverse Document Frequency refer to calculated (log(N/DF)) number of documents containing term as shown below [18] [21].

$$\mathrm{IDF} = \log \frac{D}{n_d} \qquad (2)$$

D represent sum of documents in the corpus, *nd* number of document. Hence, TF-IDF calculated as:

$$\text{TF-IDF} = \text{TF} * \text{IDF} \qquad (3)$$

Yusheng Zhou with other researchers, has implemented TF-IDF algorithm to investigate helpfulness of online reviews and which later explores correlation between review title and content on review helpfulness [21]. Other than that, in social media such as Twitter and Facebook also utilize TF-IDF method to extract and categorize sentiment in reviews from unstructured data [20] [23]. Those studies indeed help decision making by identify polarity (positive, neutral, negative) of reviews using sentiment analysis method. The researcher also capitalizes TF-IDF extraction method with Amazon dataset to analysis customers' reviews on products in making decisions and improving performance of retailers [24]. Other than English language reviews, Bahasa reviews from Tokopedia website been collected to analysis and identify hidden pattern which supports companies using TF-IDF method using sentiment analysis process [17].

B. Phrase Based

Phrases based filtration rarely been implemented by researcher due low frequency, noisy appearance with synonym words and second class statistical properties [12] [25]. The advantage of phrase-based method would be less doubtful and more presentable and represents accurate result on phrase basis.

C. Concept Based

This method concerns more on relevant and extract valuable meaning of sentences using natural language processing [18] [12] [18] [25]. Concept based model consist of three components. The first component is to analyze synonym structure of sentences by extracting verbs and arguments from text and then, to present Conceptual Ontological Graph (COG) model as one to one relationship among constituents [12] [26]. The third is to extract information using vector space model whereby it helps screening importance term in every sentences. Concept based mining measure closeness between the documents to evaluate usefulness of concept sentence level (Conceptual term frequency, ctf), document level (term frequency, tf) and corpus level (Document frequency, df).

D. Pattern Based

In text document reviews, pattern based method uses pattern deploying and pattern evolving for discover hidden pattern and trend [27]. Analysis on pattern based is discovered with method such as association rule, frequent item set mining, sequential and closed pattern mining [12]. This helps to reduces low frequency and misinterpretation by leading performance and support of related patterns. Pattern based techniques include algorithms such as Generalized Sequential Patterns (GSP), Prefix-Projected Sequential Pattern Mining (PrefixSpan), Suffix Arrays, Sequence Joining and nGram Linking [27]. Based on investigation, evaluation of those algorithms to Sequence Joining give preferably best result compared to others algorithm. The researcher works with pattern based method using association rules to determined hidden trend and sentiment in online text from social network [28].

IV. Data and Methods (Methodology)

This research propose sentiment analysis techniques and term based method to identity polarity of reviews from E-commerce site whereby the customer reviews has two components, namely, ratings and reviews [5] [7] [9]. These two columns have been incorporated in this study for assessment. Fig. 1 shows methodology of sentiment analysis to be develop using Amazons' electronic customer reviews whereby the analysis process begin with data pre-processing (refer section B), than feature extraction (refer section C), sentiment classification (refer section D) and finally Evaluation score (refer section E). The method develops using python Jupiter notebook, Anaconda.Navigator 1.10.0 whereby it is free open-source distribution can be supported in windows, macOS and Linux. SciKit-Learn and Natural Language Tool Kit (NLTK) libraries used for develop this model.

A. Data Collection

The compiled data for this paper are extracted from Amazon electronic category; it is downloaded from Kaggle in English language [29]. The file was in Comma Separated Values (CSV) whereby it's convenient to consume in python. The sample data provide customers' rating of score 1 to 5 stars and reviews which was written in English language and covers electronic products which been purchased online using Amazon.com. In total, 34,633 customers' reviews were collected from the website. For each review, rating and product details are provided. The variable and description of dataset are shown in Tables I and II. Table II provides number reviews by classes from 1 to 5. The dataset has been chosen to identify polarity of reviews.

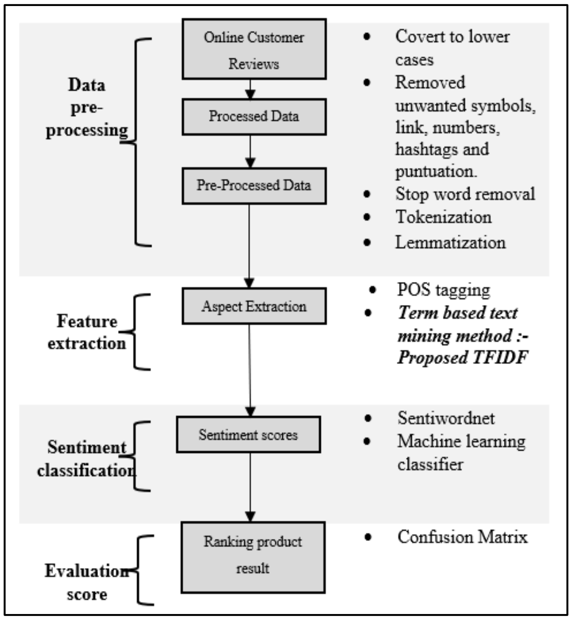


Fig. 1. Sentiment Analysis Process Flow.

TABLE I. Description of Dataset

No	Variables	Description of variables	Variables choose for this study (Yes/No)	Description
1	Id	Products' ID	No	N/A
2	Name	Products' Name	No	N/A
3	Asins	Amazon standard identification numbers	No	N/A
4	Brand	Products' brand	No	N/A
5	Categories	Products' categories	No	N/A
6	Manufacturer	Amazon or Amazon Digital Services	No	N/A
7	Review Date	Date of review data	No	N/A
8	Review ID	Unique ID for review	No	N/A
9	Review Rating	Customers' rating	Yes	Dependent Variable :- The rating is given by customer to the purchase product in the form of score ranging from 1 to 5.
10	Reviewer Name	Name of the reviewer	No	N/A
11	Review Title	Title of the review	No	N/A
12	Review Text	Text of the review	Yes	Independent Variable:- The reviews is unstructured text where expressed sentiments on purchase product positive or negative.

TABLE II. CONTENT OF DATASET

Number of Reviews	34660
Number of Columns	21
Number of reviews with rating 5	23774 (68.66%)
Number of reviews with rating 4	8541 (24.67%)
Number of reviews with rating 3	1499 (4.33%)
Number of reviews with rating 2	410 (1.16%)
Number of reviews with rating 1	402 (1.18%)

B. *Data Preprocessing*

Pre-processing phase is very important method in sentiment analysis for present quality result and enhance accuracy of the classifier to customers, it applies to Amazon dataset as shown in Table I. It also converts unstructured data to structured data which is suitable format for feature extraction. The first step of data preprocessing is converting selected alphabets into lower cases and omit all unwanted symbols, links, numbers, hashtags and punctuation. Followed by, step as below:

Convert to lowercase: The model will consider upper case as different words, hence converting to lowercase will remove noise in dataset.

Removal of unwanted symbols, links, numbers, hashtags and punctuation: By removing those details can reduce the feature space and which does not help in performance of result.

Stop word removal: Stop words will ~~is~~ not required for analysis, hence, it been removed from reviews for simplify of the text and improve performance of result. Example of stop words are ‘a’, ‘is’, ‘are’ and ‘that’.

Tokenization: The process of breaking the sentences into phrases and words.

Lemmatization: Method for switching the words into root meaning. For example, ‘used’ to ‘use’.

POS tagging: Part-of-speech tagging where it will identify each words of reviews from noun, adjective, verb and adverb.

After preprocessing, feature extraction is takes place for sentiment analyses.

C. *Aspect Extraction*

Part-of-Tagging (POS) is preferred method to identify each words of review noun, adjective, verb and adverb [15] [30]. Table III and Table IV show sample of POS tagging results after preprocessing.

Based on word segmentation, the keywords from text been identified for further process. Then, term-based text mining method was performed to generate most frequent itemset with TF-IDF model. In between, N-gram play important role to represents number of texts as unigram (1), bigram (2) and trigram (3) [16] [20]. When applying N-gram it can represent numbers of word needed in frequent itemset. Fig. 2, 3 and 4 show comparative evaluation and analysis on most frequent words based on N-gram features. The summary result present that ‘great’ is most frequent unigram words used in Amazon product dataset after removing all noisy or unwanted words whereas based on Fig. 3 ‘easy use’ is most frequent bigram words and Fig. 4 ‘amazon fire tv’ show most frequent trigram words in selected dataset.

TABLE III. SAMPLE RESULT -1 FROM DATA PRE-PROCESSING

Original Review	great for beginner or experienced person. Bought as a gift and she loves it
After convert to lower cases remove unwanted symbols	great for beginner or experienced person bought as a gift and she loves it
After removal of stop words	great beginner experienced person bought gift loves
After Lemmatization	['great', 'beginner', 'experienced', 'person', 'bought', 'gift', 'loves']

TABLE IV. SAMPLE RESULT-2 FROM DATA PRE-PROCESSING

Original Review	great for beginner or experienced person. Bought as a gift and she loves it
Star Rating	5
After POS tagging	[('great', 'JJ'), ('beginner', 'NN'), ('experienced', 'VBD'), ('person', 'NN'), ('bought', 'VBD'), ('gift', 'NN'), ('loves', 'NNS')]

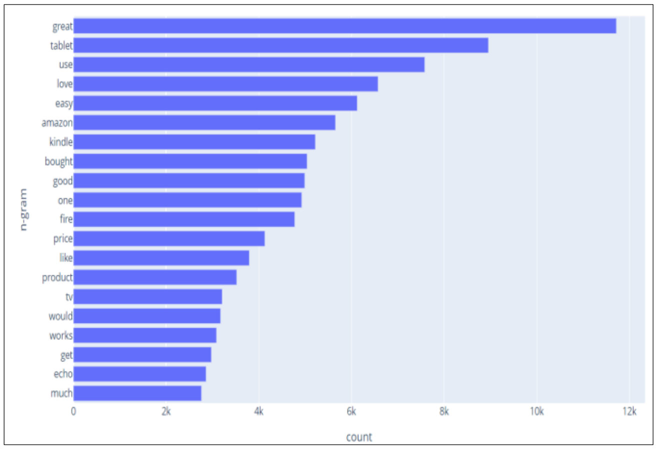


Fig. 2. Most Frequent Unigram Words.

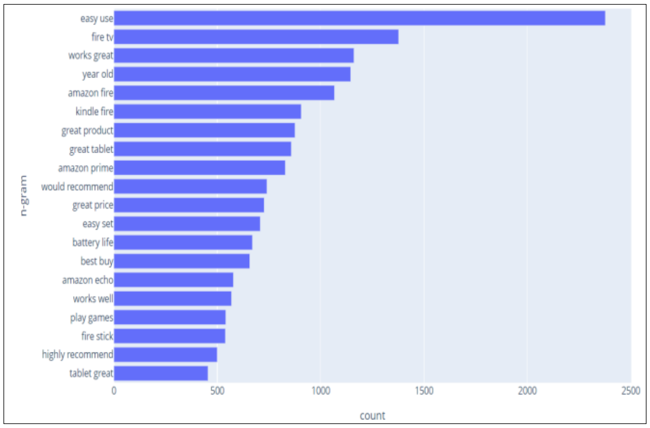


Fig. 3. Most Frequent Bigram Words.

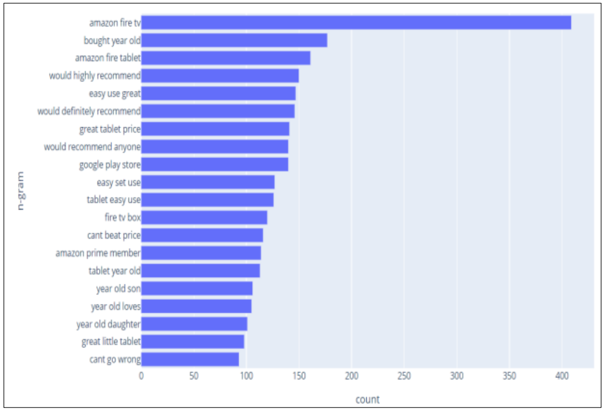


Fig. 4. Most Frequent Trigram Words.

D. Sentiment Classification

Lexicon based method SentiWordNet is used for expressed sentiment in words by scoring each words [31] [32] [33] [34]. SentiWordNet is like dictionary assign to each synset of WordNet with English language with positive and negative scores. Over 100,000 English words are in SentiWordNet for sentiment approaches with positive and negative scores. Based on POS tagging, for each synset are assign to scores from 0 to 1. The final positive and negative score of each sentence calculated with below equation:

$$\text{pos\_score} = \sum_{i=i}^{n} pos\_score\_senti(i) \tag{4}$$

$$\text{neg\_score} = \sum_{i=i}^{n} neg\_score\_senti(i) \tag{5}$$

$$\text{senti\_score} = \begin{cases} \frac{pos\_score}{tota\_score} if\ pos > neg \\ \frac{neg\_score}{total\_score} if\ neg > pos \end{cases} \tag{6}$$

Total senti score below and equal to 0 is consider negative reviews and above 0 is consider as positive reviews. Based on Table IV total score of review is 0.875 which consider as positive review. In line with this finding, diverse machine learning classification methods been used to evaluate sentiment on reviews. First of all, training and testing samples are divided as trained sample 70% and test sample 30%. While four classifier Naïve Bayes, Support Vector Machine, Decision Tree and K-Nearest Neighbour are applied for find accuracy, precision, F1-score and recall. Machine learning (ML) approach is focus on future decision making to manage textual data with intelligence whereby different ML algorithms provide different result for comparison [9] [16] [35].

$$\text{Precision} = \frac{(True\ Positive)}{True\ Positive\ + False\ Positive} \tag{7}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive\ + False\ Negative} \tag{8}$$

$$F_{Score} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{9}$$

E. Evaluation Score

After comparing four machine learning approaches in Amazon reviews the result is evaluated with confusion matrix hence it present true positive (TP), false positive (FP), true negative (TN) and false negative (FN).

TP : The output where it correctly predicts the positive class.

FN : The output where incorrectly predicts the negative class and mislabeled as negative.

FP : The output incorrectly predicts the positive class whereby mislabeled as positive.

TN : The output where correctly predicts the negative class.

V. Result and Discussion

As discussed in Section III, the paper develop model on term-based sentiment analysis process with Amazon products. The textual data analyzed using python programming language and the result obtained from experiments shown in Table V. After data pre-processing, feature extraction demonstrates frequent itemset for polarize the reviews into sentiments [6] [9]. Data pre-processing first that required conducted data analysis for reduction of noise and for perform best result [5] [20]. In order detected overall sentiment of dataset 4 different type classification methods like Naïve Bayes, Support Vector Machine, Decision Tree, K-Nearest Neighbour approach with trained and tested using Amazon products dataset. Based on 4 machine learning classifier and by N-gram applied Decision tree show highest accuracy and Support Vector Machine show highest precision, recall and F1-Score. The result indicates Support Vector Machine model performs well with proposed method with Amazon product dataset. In between, unigram show us better performance whereas trigram shows us poor performance. The N-gram weight was most traditional features applied compared the best accuracy as the result present unigram perform well compared to other N-gram [16]. Hence, TF-IDF features with N-gram show different results as different machine leaning models. The approach is efficient and more robust for process and evaluates unstructured Amazon product reviews [17]. The similarity performance is only seen in Naïve Bayes model for precision, recall and F1-score. Table V shows average summary result for positive, negative and neutral classes for different machine leaning models with TF-IDF and N-gram.

TABLE V. Summary of Performance with TF-IDF and N-gram

	Accuracy (%)			Precision (%)			Recall (%)			F1-Score (%)		
N-Grams	1	2	3	1	2	3	1	2	3	1	2	3
Naïve Bayes	78.869	79.01	79.23	63	63	63	79	79	79	70	70	70
Support Vector Machine	82.27	79.22	79.04	82	79	67	80	79	79	72	70	70
Decision Tree	99.98	99.91	99.59	78	73	74	77	64	39	78	67	45
K-Nearest Neighbour	63.33	33.83	58.75	75	66	70	50	19	29	56	19	33

VI. Contribution

Online ratings and reviews evaluate customer view and influence sales performance. Another important contribution is had been information show attitude and behavior of customers toward purchase products to make decision which led to customers' satisfaction. Sentiment detection from online reviews also influences prospective customers on online decision purchasing and a better understanding on products. The classification of reviews positive, negative and neutral evaluate subjective information and describe nature of opinion in better way. Other than that the manipulation of unstructured data being processed with intelligent technology is for present quality text information for judging human behavior. Hence, based on analyze result retailers can improve their products' price, quality and services.

VII. Conclusion

There are two main dimensions for identifying sentiments in Amazon online products with reviews and star ratings. The proposed text preprocessing and machine learning model with TF-IDF and N-gram method have been used for investigate performance of Amazon products dataset as the result as discuses in section 5. The most important insight from this study would be classification of reviews in positive, negative and neutral for present overall performance result. For future work, different text mining method might improve accuracy result for Amazon product dataset. Furthermore, different sentiment analysis techniques such as hybrid deep learning models can be proposed together with N-gram features to perform better result and comparison. In overall, we believe all features applied in this study would improve the performance of E-commerce sites.

Acknowledgment

The authors would like to thank Ministry of Higher Education (MOHE) and Universiti Teknologi Malaysia (UTM) for their educational and financial support. This work is conducted at Razak Faculty of Technology and Informatics, under Cyber Physical Systems Research Group and funded by MOHE (FRGS: R.K130000.7856.5F026) and (UTM GUP: Q.K130000.2538.18H42).

References

[1] C. C. B. T. Alexander Ligthart, "Systematic reviews in sentiment analysis: a tertiary study," Springer, p. 57, 2021.

[2] D. H. G. Tayybaha Quyyam, "Sentiment Analysis of Amazon Customer Product Reviews: A Review," IJSRED, vol. 4, no. 1, p. 32, 2021.

[3] A. S. M. AlQahtani, "Product Sentiment Anlaysis for Amazon Reviews," IJCSIT, vol. 13, p. 16, 2021.

[4] S. S. A. B. U. Rahul Ramachandrana, "Exploring the relationship between emotionality and product star ratings in online reviews," ScienceDirect, p. 10, 2022.

[5] M. J. Budhwar and P. Singh, "Sentiment Analysis based Method for Amazon Product Reviews," IJERT, 2021.

[6] T. Sinnasamy and N. N. Amir Sjarif, "A Survey on Sentiment Analysis Approaches in e-Commerce," IJACSA, vol. 12, p. 6, 2021.

[7] U. Ahmed Chauhan, M. T. Afzal, A. Shahid, M. Abdar, M. E. Basiri and X. Zhou, "A comprehensive analysis of adverb types for mining user sentiments on amazon product reviews," World Wide Web, p. 20, 2020.

[8] Y. A. Amrani, M. Lazaar and K. E. El Kadiri, "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis," ScienceDirect, p. 10, 2018.

[9] S. Wassan, X. Chen, T. Shen, M. Waqar and N. Jhanjhi, "Amazon Product Sentiment Analysis using Machine Learning Techniques," REVISTA ARGENTINA, p. 9, 2021.

[10] D. A. B. Tanvi Hardeniya, "Dictionary Based Approach to Sentiment Analysis - A Review," IJAEMS, p. 6, 2016.

[11] P. Carracedo, R. Puertas and L. Marti, "Research lines on the impact of the COVID-19 pandemic on business. A text mining analysis," ScienceDirect, p. 8, 2020.

[12] A. Raut and P. V. Shinde, "Effective Methods and Techniques in Text Mining," IJRITCC, vol. 5, no. 3, p. 4, 2017.

[13] H. S.Choi and S. Leon, "An Empirical Investigation of Online Review Helpfulness: A Big Data Perpective," ScienceDirect, p. 12, 2020.

[14] M. Ahmad, S. Aftab and N. Hameed, "Sentiment Analysis using SVM : A Systematic Literature Review," IJACSA, vol. 9, p. 7, 2018.

[15] M. P Abraham and U. K. Reddy, " Feature Based Sentiment Analysis of Mobile Product Reviews using Machine Learning Techniques," IJATCSE, vol. 9, p. 8, 2020.

[16] D. Gamal, M. Alfonse, E.-S. M. E1-Horbaty and A.-B. M.Salem, "Implementation of Machine Learning Algorithms in Arabic Sentiment Analysis Using N-Gram Features," ScienceDirect, p. 9, 2019.

[17] P. Kencana Sari, A. Alamsyah and S. Wibowo, "Measuring e-Commerce service quality from online customer review using sentiment analysis," IOP, p. 7, 2018.

[18] S. D. K. and D. S. , "The Comparison of Term Based Methods Using Text Mining," IJCSMC, vol. 5, no. 9, p. 5, 2016.

[19] A. Masood Khan and K. Rahat Afreen, "An approach to text analytics and text mining in multilingual natural language processing," ScienceDirect, p. 3, 2020.

[20] R. Ahuja, A. Chug, S. Kohli, S. Gupta and P. Ahuja, "The Impact of Features Extraction on the Sentiment Analysis," ScienceDirect, p. 7, 2019.

[21] Y. zhou, S. Yang, Y. Li, Y. chen and J. Yao, "Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining," ScienceDirect, p. 11, 2020.

[22] M. Mujahid, E. Lee, F. Rustam, P. B. Washington, S. Ullah, A. Ahmad Reshi and I. Ashraf, "Sentiment Analysis and Topic Modeling on Tweets about Online Education during COVID-19," MDPI, p. 25, 2021.

[23] S. T. Y. L. H. S. Z. G. X. Y. J. B. C. C. and X. H. , "Interpreting the Public Sentiment Variations on Twitter," IEEE, vol. 26, p. 13, 2014.

[24] V. Raghuraman, S. Pattanayak, A. VK and A. C Patil, "Sentiment Analysis on Amazon Product Reviews," IEEE, p. 6, 2020.

[25] R. and S. , "A Study Of Text Mining Methods, Applications, and Techniques," IJESRT, p. 6, 2017.

[26] S. Hassan, F. Karray and M. Kamel, "Concept Mining using Conceptual Ontological Graph (COG)," ResearchGate, p. 11, 2014.

[27] P. Ozdzynski and D. Zakrzewska, "Using Frequent Pattern Mining Algorithms in Text Analysis," Information systems in Management, p. 10, 2017.

[28] J. A. Diaz-Garcia, M. Ruiz and M. J.Martin Bautista, "Non-Query-Based Pattern Mining and Sentiment Analysis for Massive Microblogging Online Texts," IEEE, p. 17, 2020.

[29] Datafiniti, "Kaggle," kaggle, 2020. [Online]. Available: https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products. [Accessed 2022].

[30] S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," IEEE, p. 5, 2018.

[31] S. K. Kochhar and U. Ojha, "Index for objective measurement of a research paper based on sentiment analysis," ScienceDirect, p. 7, 2020.

[32] S. A. Firmanto and R. Sarno, "Prediction of Movie Sentiment based on Reviews and Score on Rotten Tomatoes using SentiWordnet," IEEE, p. 28, 2018.

[33] N. M. Elfajr and R. Sarno, "Sentiment Analysis using Weighted Emoticons and SentiWordNet for Indonesian Language," IEEE, p. 26, 2018.

[34] P. Mehta and S. Chandra, "Parameter Tuning in Updating the Sentiment Polarity of Objective Words in SentiWordNet," IEEE, p. 43, 2015.

[35] N. Nandal, R. Tanwar and J. Pruthi, "Machine learning based aspect level sentiment analysis for Amazon products," ResearchGate, p. 8, 2020.

An Iootfuzzer Method for Vulnerability Mining of Internet of Things Devices based on Binary Source Code and Feedback Fuzzing

Guangxin Guo, Chao Wang*, Jiahan Dong, Bowen Li, Xiaohu Wang
State Grid Beijing Electric Power Research Institute, Beijing, China

***Abstract*—With the technological progress of the Internet and 5G communication network, more and more Internet of Things devices are used in it. Limited by the cost, power consumption and other factors of Internet of Things devices, the systems carried by the Internet of Things devices often lack the security protection provided by larger equipment systems such as desktop computers. Because the current personal computers and servers mostly use the x86 architecture, and the previous research on security tools or hardware-based security analysis feature support is mostly based on the x86 architecture, the traditional security analysis techniques cannot be applied to the current large-scale ARM-based and MIPS-based Internet of Things devices. Based on this, this paper studies the firmware binary program of common Linux-based Internet of Things devices. A binary static instrumentation technology based on taint information analysis is proposed. The paper also analyzes how to use the binary static instrumentation technology combined with static analysis results to rewrite binary programs and obtain taint path information when binary programs are executed. Firmware binary fuzzing technology based on model constraints and path feedback is studied to cover more dangerous execution paths in the target program. Finally, iootfuzzer, a prototype vulnerability mining system for firmware binaries of Internet of Things devices, is used to test and analyze the two technologies. The results show that its fuzzing efficiency for Internet of Things devices is better than other fuzzing technologies such as boofuzz and Peach 3. It can fill in some gaps in the current security analysis tools for the Internet of Things devices and improve the efficiency of security analysis for Internet of Things devices, which contributes to the field through automated security vulnerability detection systems.**

Keywords*—Internet of things; system vulnerabilities; source code; fuzz testing; instrumentation technology

I. Introduction

At present, the interaction between IoT devices and the outside world is mostly carried out through the network. Usually, the software monitors the input data of external users, and the user's operation on the device is received and processed through several specific softwares [1]. Therefore, to analyze the vulnerability of this software, it is necessary to start from the code path through which external user data flows and find the problem code that external users, or attackers, can reach [2].

At the same time, in the past commonly used fuzzy testing based on path feedback, almost all of them adopt the way of the full instrumentation, such as path record instrumentation for all jump instructions at the end of the code block [3]. This instrumentation method will lead to a lot of programs internal processing codes unrelated to user input being instrumented, and then will make the code unrelated to external input data generate a new execution path so that the fuzz test tool will mistakenly think that the new execution path is caused by the fuzz test sample, and use the fuzzing test sample to further mutate and test in the follow-up. This will reduce the efficiency of fuzz testing.

At present, the work of source code vulnerability mining mostly depends on manually defined rules, but this method has some drawbacks: on the one hand, manually defined vulnerability mining rules often need to rely on the expertise and work experience of experts [4]. It is difficult to ensure full coverage of the possible causes of vulnerabilities, and it is easy to have the possibility of false positives and underreporting. On the other hand, it takes a lot of manpower to mine loopholes according to the rules. Due to manual judgment, the phenomenon of false underreporting will still occur [5]. With the development of technology, researchers began to use machine learning methods for vulnerability mining. This method does not need to define vulnerability rules, but still needs to define vulnerability features. Although it has reduced the manual workload, there are still drawbacks to feature coverage. In recent years, with the increasing popularity of deep learning, many scholars have begun to try to apply deep learning methods to vulnerability mining, and have made good progress [6]. In the process of data preprocessing, only whether the source code contains vulnerabilities is divided, and the types of vulnerabilities are not classified, so the existing work can only detect whether a piece of code contains vulnerabilities, and cannot accurately detect the types of vulnerabilities.

Scandariato et al. [7] tested the bag-of-words technique with a hybrid approach combining N-gram parsing and statistical feature selection to predict vulnerable software components. Yamaguchi [8] proposed a new graphical representation, called the code property graph, by traversing the graph to discover vulnerabilities. However, designing effective traversals to detect complex vulnerabilities can be very difficult. Thus, the authors propose an automatic method for traversing code attribute graphs to effectively locate taint-style vulnerabilities generated by uncleaned data streams, and use it to experiment with five popular open-source projects. The number of source codes without vulnerabilities in the data set is much larger than the number of source codes with vulnerabilities. It is difficult to learn the characteristics of a

*Corresponding Author.

small number of samples, while a large number of samples are prone to over-fitting during training.

To solve the problem that it is difficult to apply the fuzzing technology based on path feedback to the firmware binary program of ARM and MIPS Internet of Things devices and the efficiency of fuzzing test is low, this paper proposes a binary static instrumentation technology based on the taint information analysis. Firmware binary fuzzing testing technology based on model constraints and path feedback is studied in depth. The binary static instrumentation technology based on taint information analysis is studied, which can solve the problem that the current ARM and MIPS architectures lack taint information analysis tools. It analyzes and instruments the conditional branch jump of the target program in firmware affected by external input data to provide information feedback to the fuzz testing tool. The firmware binary fuzzing testing technology based on model constraints and path feedback is studied, which can improve the efficiency of fuzzing testing technology such as model constraints and path feedback combined with the above technology applied to the vulnerability mining of the target device, and focus the fuzzing test on the dangerous path to achieve the effect of improving the efficiency of fuzzing testing. The abstract syntax tree of function is used to represent the function, and multiple functions are annotated in open source datasets to capture the intrinsic representation of vulnerabilities, which proves that the model is effective for cross-project vulnerability detection at the functional level. Finally, the effectiveness of the above technology is proved by comparative experiments.

The main innovations of this paper are:

1) Adopt a feedback type fuzz test technology to carry out vulnerability mine on that firmware of the Internet of things equipment, and select the conditional branch jump points influenced by external user input data, namely taint information, as path feedback data of the fuzz test.

2) Design and implement a firmware vulnerability mining prototype system based on binary static instrumentation and feedback fuzzing for ARM and MIPS architectures.

3) Based on the results of taint information analysis, the target binary program is instrumented to improve the execution efficiency of the fuzz testing process.

Through the binary static technology based on the Internet of Things device vulnerability analysis, the fuzzing testing tool can obtain the feedback information of the relevant test samples in the process of fuzzing testing, and guide the generation of the fuzzing test samples. At the same time, the relevant algorithm is designed to select more valuable samples, and the model constraints are used to make the fuzzy test towards a higher coverage.

II. Related Work

A. Binary Static Instrumentation Technique

Static Binary Instrumentation (SBI) modifies the binary program file stored in the storage medium to generate an instrumented binary file and save it to the storage medium. When the program is executed, the instrumented binary file is run [9]. The idea of binary static instrumentation technology is very simple, which is to modify the binary file through the target file format to achieve the purpose of adding specific new code, but it needs to statically analyze the instrumented program in advance or describe it through the configuration file [10]. One of the core parts of binary static instrumentation technology is the selection of instrumentation positions. An example of an instrumentation error is shown in Fig. 1.

Static instrumentation technology modifies and hijacks the original code execution flow of the program so that the program jumps to the instrumentation code to execute when it runs to a specific location and completes the specific functions inserted by developers [11]. The selection of the insertion point will not only affect the efficiency of the program execution but also affect whether the program can be executed normally. In instrumentation tools that use binary static instrumentation techniques, directly editing the target binary file format and modifying and inserting code is the most common implementation.

B. Feedback Fuzzy Test Technique

Under the condition that the source code can be instrumented, AFL, honggfuzz, and other fuzzing tools all adopt the feedback fuzzing technology. These fuzz testing tools use the execution path information of fuzz test samples as feedback information to guide the sample generation tool to generate samples that can improve the coverage of fuzz test code (path) [12]. AFL uses customized GCC to insert the functional code of path recording and path feedback into the compiled binary program through source code instrumentation so that the program can record and feedback the execution information of fuzzy test samples during running [13]. In the absence of source code, binary instrumentation is generally used to obtain the key information in the running process of the target program, which requires researchers to develop and customize it for specific situations. Under the condition of no source code, AFL can still fuzz binary programs with the QEMU tool, but the efficiency of fuzz testing is low [14]. InsFuzz uses binary static instrumentation to interpolate non-source binary programs, inserting path records and feedback code into binary programs to achieve the same effect as source instrumentation [15]. Through the feedback fuzzing testing technology, the efficiency of fuzzing testing can be effectively improved. Inefficient fuzzing test samples can be discarded in time, and only the fuzzing test samples which may generate new execution paths can be mutated so that the whole fuzz testing work can be carried out in the direction of improving the code test coverage.

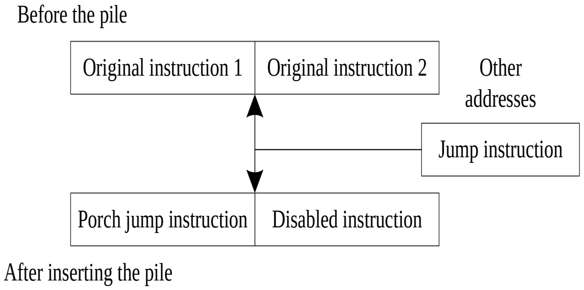


Fig. 1. Example of CISC Instrumentation Causing Program Runtime Errors.

III. Research on Binary Static Instrumentation Technology based on Stain Information Analysis

C. Analysis Flow and Algorithm of Taint Information

To improve the effective code coverage rate of the subsequent fuzz testing tool, the technology performs simulation execution on the target binary program, and therefore, more functional codes need to be analyzed as much as possible in the taint information analysis process. Because the target binary program in this paper is the network program in the firmware of the Internet of Things device, which uses socket, bind, accept, fgets, send, and other functions to build the data receiving and sending part of the network application program. These programs all have a clear function to receive external input data. The Internet of Things device firmware program uses the fgets function to receive external data packets transmitted through the network through the file descriptor of the socket.

Since program codes are executed in a simulation execution mode when taint information analysis is carried out [16], even if the analysis is not purely static, partial run-time information is missing. The coding fragments are shown in Table I.

TABLE I. Code Fragment

Code fragment:
<META HTTP-EQUIV="0,0x271t" CONTENT="no-cache"> If iggate,10000,(ffile * dword_654BC V0=400; V1=”Bad Request”; V2=”No request found” Return sub_D343（V0, V1, V2）; <META HTTP-EQUIV="Cache-Control" CONTENT="no-cache, must-revalidate">

As shown in the variable s in Table I, during simulation execution and taint analysis, it is impossible to accurately assign the external input variable or simulate the external input that can explore all execution paths [17]. Therefore, it is necessary to apply some methods to analyze the code that operates on the external input data as much as possible during simulation execution and taint analysis.

D. Research on Firmware Binary Fuzz Testing Technology for Internet of Things Devices based on Model Constraints and Path Feedback

- Fuzz testing coverage and sample selection algorithm

When the feedback information received by the sample variation module finds that an execution path is generated, it indicates that the current fuzzy test sample triggers a new execution path [18], which means that the change of some fields in the sample triggers the change of the flow direction of the program control stream, and by performing sample variation on the sample again and generating a new sample. There is a greater probability that the code coverage of the fuzz test can be improved, so it is placed in the queue of samples to be mutated.

In the whole process of fuzzing testing, this paper defines an index of dangerous branch jump coverage, which is used to record and judge the index data in the process of fuzzing testing:

$$C_{danger} = \frac{DB_{executed}}{DB_{all}} \times 100\% \tag{1}$$

In Formula 1, DB_{all} represents the number of branch jumps influenced by the external user input data in the target binary program, and $DB_{executed}$ represents the number of branch jumps influenced by the external user input data that have been executed in the current fuzz test.

A queue $Q = \{S_0, S_1, \cdots, S_n\}$ and a queue $M = \{m_0, m_1, \cdots, m_n\}$ of fuzzing test samples to be tested are defined for the fuzzing test samples transmitted to the fuzzing testing module. During the fuze testing process, the fuze test samples in the queue Q are transmitted to the fuzz testing module in a first-in first-out order [19]. If a new execution path is generated after the fuzz test samples are executed, then pass it to the use case generation function and add the newly generated fuzzing test sample to the fuzzing test sample queue M.

- Target feedback data processing

An array $DB = \{DB_0, DB_1, \cdots, DB_n\}$ is used to record the execution of the current fuzz test sample in the target binary program. This data is fed back after the target binary program executes the external input data processing function. At the same time, the array $AC = \{AC_0, AC_1, \cdots, AC_n\}$ is defined to record the total number of times that the branch jump of each taint condition is executed in the fuzz test so far.

As shown in Equation 2, a weight must be defined for each fuzzy test sample to represent the "value" of the fuzzy test sample:

$$\begin{cases} P_{ij} = 1 \quad (DB_{ij} \neq 0) \\ P_{ij} = 0 \quad (DB_{ij} = 0) \\ W_i = \sum_{j=0}^{n} \left(\frac{P_{ij}}{AC_j} \times 1000 \right) \end{cases} \tag{2}$$

The weight of each branch is $1000 / AC_i$. When a taint condition branch jumps in the fuzz test and the total number of times it has been executed so far is AC_i, its reciprocal is multiplied by 1000 to get its weight. The factor of 1000 is used to prevent too many executions from causing the reciprocal to be too small, and the value becomes 0 after the decimal place is normalized. The factor can increase with the number of fuzz tests without affecting the weight ordering.

IV. Experimental Analysis

E. *Function Comparison of Stain Analysis Tool*

At present, there is no mainstream tool of the same type supporting ARM and MIPS that is open source or available for download for comparison, as shown in Table II.

TABLE II. Comparison Results of Stain Analysis Tools

Tool name	x86	arm	mips	Required documents
Pin	Y	N	N	Source Code
Taintgrind	Y	Y	N	Source Code
TaintEraser	Y	N	N	Source Code
TEMU	Y	N	N	Source Code
PyDaint	Reserved interface	Y	Y	Binary

Most mainstream tools rely on the source code to recompile the analysis target before running the analysis. PyDaint not only supports the ARM and MIPS architectures studied in this paper, but also can be extended to further support x86 architectures [20].

The test framework is shown in Fig. 2.

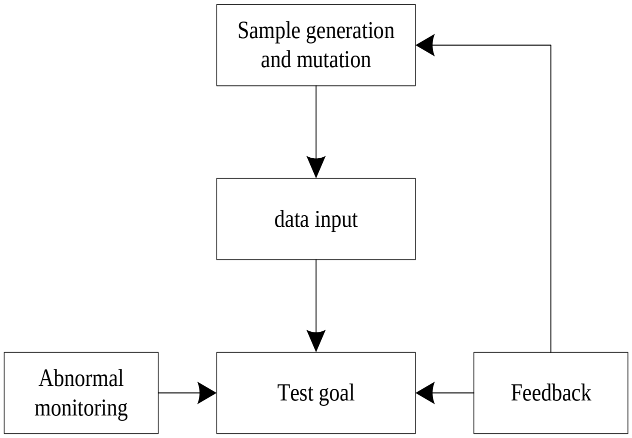


Fig. 2. Basic Framework of Paste Test.

F. *Instrumentation Performance Test of Tainted Information Flow*

Because this project uses QEMU user mode to run simulation on x86 _ 64 computers for ARM and MIPS architectures, it is difficult to carry out relevant timing statistics. Therefore, to test the effect of binary static instrumentation based on taint information analysis on the execution efficiency of the original binary program, the Web service programs of ASUS AC88U router based on ARM architecture and D-LinkDIR882 router based on MIPS architecture are instrumented respectively. It takes for normal user interaction (simulated access and packet reception through a Python program) before and after instrumentation in a test environment. The experimental results are shown in Table III and Table IV.

TABLE III. Comparison of Execution Efficiency before and after HTTPD Instrumentation in ARM Architecture

Test object Sample number	Before insertion	After insertion
100	432.58ms	846.39ms
1000	4545.74ms	8682.94ms

The average time for httpd to send and receive data is about 4.3ms before instrumentation, and 8.7ms after instrumentation. In the experimental environment, the instrumentation loses about 1.02 times of performance.

TABLE IV. Comparison of Execution Efficiency of Lighttpd before and after Instrumentation of MIPS Architecture

Test object Sample number	Before insertion	After insertion
100	32.59ms	96.34ms
1000	228.39ms	972.21ms

The average time for lighttpd to send and receive data is about 0.28ms before instrumentation, and 0.99ms after instrumentation. In the experimental environment, the instrumentation loses about 2.5 times of performance.

G. *Fuzzy Test Function Comparison*

Before analyzing the experimental results of fuzz testing efficiency, as shown in Table V, we first compare the functions of several common fuzz testing tools with the fuzz testing tool iboofuzzer implemented in this paper.

TABLE V. Comparison Results of Fuzz Test Tools

Tool names	Network Program Fuzziness Test	Path feedback	Path feedback without source code	Model constraints
AFL	N	Y	Partially supported	N
AFL-net	Y	Y	Partially supported	N
Peach 3	Y	N	N	Y
boofuzz	Y	N	N	Y
iboofuzzer	Y	Y	Y	Y

Iboofuzzer is a fuzzy testing subsystem developed for the network binary program in the Internet of Things device firmware in this paper, which supports model constraints and path feedback, and can obtain the feedback information of the target binary program by using instrumentation tools in common use scenarios without source code.

H. *Fuzzy Test Efficiency Test*

When using the original boofuzz for fuzz testing, the sample random mutation function is added to avoid the premature end of the fuzz test due to sample exhaustion. At the same time, the iboofuzzer framework is used to count the coverage information of boofuzz and Peach 3 tests, but it is not fed back to the sample generation module for analysis and comparison. In the testing process, the original boofuzz is tested based on model constraints and random data. For Peach

3, its listening mode is used, and the fuzzy test samples are actively obtained from Peach through the framework of iboofuzzer and then sent to the target binary program. The test parameters are shown in Table VI.

TABLE VI. FUZZY TEST TOOL EXPERIMENTAL VARIABLES

Tool names	Sample generation method	Information feedback
boofuzz	Random data	No
boofuzz	Model constraints	No
Peach 3	Model constraints	No
iboofuzzer	Model constraints	Path feedback

For each of the four test conditions in Table VI, a blur test was performed for about 8 hours.

- Comparison of coverage rate of insertion point in fuzzy test

The coverage rate of instrumentation points in the fuzzing test process represents the proportion of different branches of instrumented points executed in the whole fuzzing test process, as shown in Fig. 3.

In this comparative experiment, the coverage of boofuzz, which uses random data generation for fuzz testing, is the lowest in the whole 8-hour test. The second-lowest is boofuzz, which uses model constraints to generate samples. It is lower than Peach 3 in the first few hours, and then gradually approaches. Peach 3 has the second-highest coverage and iboofuzzer has the highest coverage. That is to say, the quality of samples generated by iboofuzzer is higher, and the test coverage of dangerous paths is larger. In the experimental environment, the coverage rate of the algorithm in this paper is the highest, and the target binary program after interpolation can still process a single fuzzy test sample in milliseconds.

- Comparison of the number of new execution paths for fuzz testing

The number of new execution paths generated during fuzz testing represents the ability of the fuzz testing tool to find new code test paths throughout the fuzz testing process, which is shown in Fig. 4.

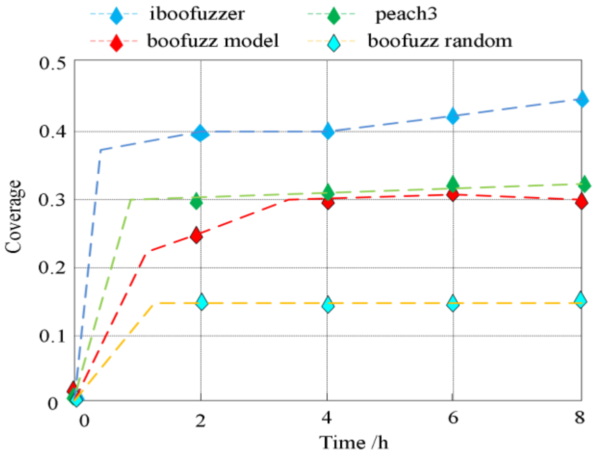


Fig. 3. Comparison of Fuzz Test Instrumentation Point Coverage.

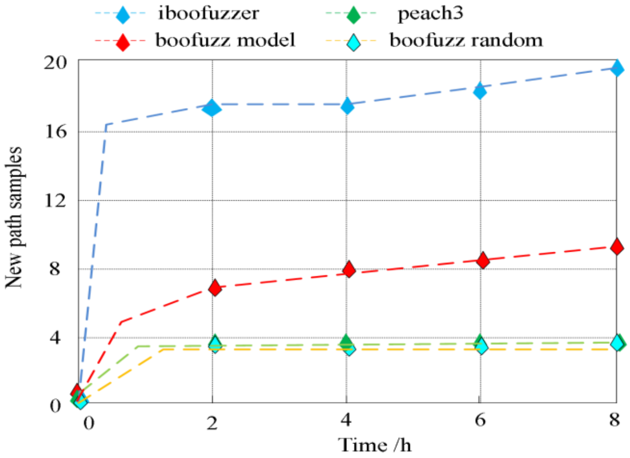


Fig. 4. Comparison Chart of the Number of New Execution Paths for Fuzz Testing.

Peach 3 and boofuzz with random data both generate fewer new path samples, while iboofuzzer generates the newest path samples, that is, the samples generated by iboofuzzer are more effective in discovering new paths. The distortion efficiency is improved by increasing the proportion of the effective distortion in the total distortion. Samples with effective distortion can reach the target basic block faster, while samples with invalid distortion will make the program fall into the situation of path explosion.

The object of vulnerability mining in this paper is the network binary program in the Internet of Things devices, which has certain format requirements for the input data, so the generation of samples based on model constraints can not only solve the problem of lack of original data samples in fuzzy testing, but also improve the code penetration of samples, and avoid the samples being abandoned in the format check function of the target program. At the same time, the fuzzy test focuses on the dangerous path affected by the external input data, and mutates new fuzzy test samples to improve the coverage of the dangerous path. To improve the efficiency of fuzzy testing, the path feedback information and the calculated sample weight are used to guide the mutation generation of fuzzy testing samples.

V. CONCLUSION

In this paper, we study the static instrumentation of the firmware binary program of the Internet of Things device based on ARM and MIPS architecture, and innovatively combine the taint information analysis with the binary instrumentation and apply it to the firmware program of the Internet of Things device to improve the efficiency of fuzzing the dangerous path in the tested program of the firmware program of Internet of Things devices. The main work includes:

1) Analyze the taint information with the firmware binary program of the Internet of Things device based on ARM and MIPS architecture.

2) The feedback fuzzing technology is applied to the firmware binaries of the IoT devices based on ARM and MIPS architectures, and the target binaries are moved from the IoT devices to desktop computers for fuzzing by using QEMU open source tools and binary static instrumentation technology.

3) Select some of the common security tools to compare with the subsystems of the prototype system iootfuzzer in different types, and use the tools that can fuzze the research goal of this paper to carry out the comparison experiment of fuzzing efficiency.

The fuzz testing subsystem in this paper is implemented based on boofbzz, and it needs to write samples to generate template files for different test objectives, which is a heavy workload. In the future, we can consider implementing the technology of automatically generating the corresponding template through the captured network communication packets to reduce the cost of preparation work in the early stage of fuzz testing. Complex vulnerabilities usually have a long ROC chain. These vulnerabilities may not only be related to one file, but also to multiple files, which will greatly affect the effect of vulnerability mining. Dynamic analysis is to detect vulnerabilities in the process of program running, which makes dynamic analysis more complex. In the future, the above two directions will be studied in depth.

REFERENCES

[1] Hoa Khanh Dam, Truyen Tran, Trang Pham, Shien Wfee Ng, Jolin Grundy, and Aditya Ghose. Automatic feature learning for vulnerability prediction. arXiv preprint arXiv: 1708.02368, 2017.

[2] Zhen Li, Deqing Zou, Shouhuai Xu, Xinyu Out Hai Jin, Sujuan Wang, Zhijun Deng, and Yuyi Zhong. Vuldeepecker: A deep learning-based system for vulnerability detection. arXiv preprint arXiv: 1801.01681, 2018.

[3] Siqi Ma, Ferdian Thung, David Lo, Cong Sun, and Robert H. Deng. Vurle: Automatic vulnerability detection and repair by learning from examples. In European Symposium on Research in Computer Security, pages 229-246, 2017.

[4] Rebecca Russell, Louis Kim, Lei Hamilton, Tomo Lazovich, Jacob Harer, Onur Ozdemir, Paul Ellingwood, and Marc McConley. Automated vulnerability detection in source code using deep representation learning. In 2018 17th IEEE49International Conference on Machine Learning and Applications (ICMLA), pages 757-762, 2018.

[5] Miltiadis Allamanis. Earl T. Barr, Premkumai Devanbu, and Charles Sutton. A survey of machine learning for big code and naturalness. ACM Computing Surveys (CSUR), 51(4):81, 2018.

[6] David Evans. Splint-secure programming lint. Technical report, Teclmical report, 2002. David Wheeler. Flawfinder home page, 2006, 2019-09-27.

[7] Seyed Mohammad Ghaffarian and Hamid Reza Shahriari. Software vulnerability analysis and discovery using machine-learning and data-mining techniques: A survey. ACM Computing Surveys (CSUR), 50(4):56, 2017.

[8] Reudismam Rolim, Gustavo Soares, Ro hit Gheyi, Titus Barik, and Loris D'Antoni. Learning quick fixes from code repositories. arXiv preprint arXiv: 1803.03806, 2018.

[9] Zheng Y, Davanian A, Yin H, et al. FIRM-AFL: high-throughput greybox fuzzing of IoT firmware via augmented process emulation[C]//28th {USENDC} Security Symposium ({USENIX} Security 19). 2019: 1099-1114.

[10] Fowler D S, Bryans J, Shaikh S A, et al. Fuzz testing for automotive cyber security[C]//2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W). IEEE, 2018: 239-246.

[11] David Liu, Andrew Petersen. Static Analyses in Python Programming Courses. 2019:666-671.

[12] Zhao Na. Research on the role of new multilateral development financial institutions in the construction of the “Belt and Road”. Journal of Economic Research. 2020(10).

[13] Zhang Fengyang. Difficulties and Suggestions of Financial Consumer Protection in Rural Financial Institutions. Business News. 2020(12).

[14] Zhu Jie, You Xiong, Xia Qing. Space-time data organization model of battlefield environment objects based on mission process. Journal of Wuhan University (Information Science Edition). 2018(11).

[15] Roberto Baldoni, Emilio Coppa, Daniele Cono D’elia, et al. A Survey of Symbolic Execution Techniques. ACM Computing Surveys, 2018, 51(3):1-39.

[16] Caroline Lemieux, Koushik Sen. FairFuzz: a targeted mutation strategy for increasing greybox fuzz testing coverage. 2018:475-485.

[17] M. A. Klimushenkova, P. M. Dovgalyuk. Improving the performance of reverse debugging. Programming and Computer Software, 2017, 43(1):60-66.

[18] Seyed Mohammad Ghaffarian, Hamid Reza Shahriari. Software Vulnerability Analysis and Discovery Using Machine-Learning and Data-Mining Techniques. ACM Computing Surveys (CSUR), 2017, 50(4).

[19] Philipp Dominik Schubert, Ben Hermann, Eric Bodden. PhASAR: An Inter-procedural Static Analysis Framework for C/C++. 2019, 11428:393-410.

[20] Shuai Wang, Dinghao Wu. In-memory fuzzing for binary code similarity analysis. 2017:319-330.

Mobile Learning in Science Education to Improve Higher-Order Thinking Skills (HOTS) and Communication Skills: A Systematic Review

Adilah Afikah[1], Sri Rejeki Dwi Astuti[2], Suyanta Suyanta[3], Jumadi Jumadi[4], Eli Rohaeti[5]
Department of Science Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia[1, 4]
Department of Chemistry Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia[2, 3, 5]

***Abstract*—Today, the increasing use of technology and mobile applications in education was interesting. This research was a systematic review study with limited 30 articles from 2012 to 2021. It aims to answer research questions about what mobile devices used in learning and what learning approaches are used in science learning to improve higher-order thinking skill and communication skills. The findings of this study were in line with the research objectives: First, the most appropriate mobile devices used to achieve learning objectives are mobile phones, followed by PDAs, tablets, iPad, laptops, e-books, and iPods. Second, the learning approach used in science learning to improve higher-order thinking skills and communication skills are a collaborative learning approach, inquiry learning, project-based learning, problem-based learning, game-based learning, and flipped classroom learning. It was hoped that this research can be an illustration for other researchers to create innovative learning approaches. Some research that can be done next based on this research is how mobile learning in social learning or comparing the two, further research on the most appropriate learning media for mobile learning, and research on the effectiveness of implementing approach strategies in mobile learning.**

Keywords—Communication skills; higher-order thinking skills; mobile learning; science education

I. Introduction

Mobile technology provides many advantages for users, such as ease of carrying, sensitivity in operating systems and applications, connecting between different spaces and times, and facilitating social interaction. The improvement of technology and mobile applications in education is a common phenomenon observed worldwide [1]. Mobile technology has great appeal for researchers, including researchers from education field [2]. Technology in education makes students enter into social learning contexts and appropriate situations with circumstances to encourage students to learn together or collaboratively.

Mobile learning is electronic learning through mobile technologies such as computers, laptops, cell phones, audio players, and electronic books [3]. Mobile learning allows students to learn in collaborative learning with other students to share ideas with the help of the Internet and technological developments that can be done without the boundaries of space and time [4]. Always-connected mobile device allow students to access course information and provide students with opportunities to interact with content and to explore it.

Mobile learning can help remove the barriers between learning and real life. Mobile learning is powerful in its appeal of bringing new space, time, and geography into the classroom. According to some researchers, mobile learning can be a bridge between formal and informal learning to judge the difference between the two based on the context and its features [5]. Most of the literature discusses the affordability of mobile learning to implement its methods, strategies, and applications, which are mostly teacher-centered.

However, the realization of its practical implementation has not been much because its implementation has various obstacles and challenges. There are limited financial resources, inadequate educational policies for mobile learning that have not developed rapidly, human resources that do not have a good understanding, and there is still a lack of skilled personnel for effective implementation of pedagogy. Another challenge is changing the pedagogical understanding that teachers have to understand mobile learning, the lack of resources for complex tools such as infrastructure and bandwidth, parental trust because there are perceptions of health and psychological problems that follow and are associated with prolonged use of mobile devices by students. And then the lack of educators trained with mobile learning [6]–[12].

In addition to the actual implementation problems that arise in learning practices, there are fundamental problems regarding the theoretical and pedagogical foundations in mobile learning applications in the world of education, which are still lacking and not yet on target. Many authors have discussed and have investigated this with a socio-constructivist approach and found that communication features on mobile devices can encourage collaboration and become the basis of mobile learning. However, students can participate and can collaborate in learning with themselves and others, not with the devices or media used [13]. Therefore, every element of education, government, teachers, students, and society, must adapt to mobile learning so that its effectiveness can be felt.

In the 21st century, one of the biggest challenges faced is maintaining students' interest and involvement to keep interacted and connected through mobile devices in effective learning [14]. The importance of developing students' higher-order thinking competencies, such as problem-solving and critical thinking skills [15]. Mobile learning can be an innovative option for education [5]. Mobile learning not only helps students and teachers understand the content of the material in learning, but also facilitate communication, problem-solving, creativity, and students' higher-order thinking skills.

Based on this, appropriate pedagogical and theoretical methods are needed to assist teachers in designing mobile learning [16]. This method has a strategy to integrate mobile learning into the classroom to achieve learning objectives with mobile devices. In addition, existing strategies must instill critical inquiry skills in learning while still presenting problems for students to provide solutions [17].

This review, derived from a systematic content analysis science education empirical research articles, sought to address the following questions:

- What mobile devices use in learning?
- What learning approaches are used in science learning to improve higher-order thinking and communication skills?

II. Literature Review

A tool containing several solutions to educational problems often describes technology [17]. Education aims to make people knowledgeable, creative, informative, digitally capable, and adaptable [5]. Furthermore, information technology such as the internet and multimedia systems that have been utilized and applied in learning aims to improve the quality of learning by facilitating students' access to adequate resources and services [18].

One of the benefits of a mobile learning system is that students can participate in different learning situations, such as school learning and online distance learning [2]. However, distance learning is not in a room like schools in general, but in a scope that is not traced by using cellular technology beyond the existing distance [19].

Mobile learning aims to provide a different and meaningful student learning experience, but it is not the primary learning method because it must have the right learning approach. Mobile devices can influence and improve student learning outcomes and motivation in learning. However, there are limitations in its application; although teachers and students will become more familiar than usual because of the efficiency and effectiveness of this method in learning, teachers and students must maintain educational ethics [3]. Research also shows that with awareness and adequate support from all aspects of education, mobile learning will answer the 21st-century challenges of learning anywhere and anytime, a centered and innovative learning method.

Higher-order thinking skills are the ability of students to think at a higher level. Students who have these abilities will analyze, evaluate, and create innovations in solving problems. These mobile and science learning skills are necessary because many problems can be solved using higher-order thinking skills [20]. Higher-order thinking skills can improve with various learning approach strategies, learning media, and teaching materials.

Cognitive taxonomy bridges understanding higher-order thinking skills' concepts and characteristics. The most popular cognitive taxonomy is Bloom's taxonomy, revised by [21], consisting of two dimensions: the dimensions of knowledge and cognitive processes. The Knowledge Dimension classifies the types of knowledge students acquire into four types: factual, conceptual, procedural, and metacognitive. While the cognitive process dimension consists of six levels: remembering (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6) [22].

The three categories of HOTS assessment abilities are as follows: first, the ability to transfer concepts to other concepts, higher-order thinking skills as a form of knowledge possessed, the ability to relate to other people in unfamiliar situations. Second, critical thinking skills are the ability to understand logical problems, reflective thinking skills, and the ability to argue that can focus on making decisions or doing something. Third, problem-solving skills (problem-solving), namely the ability to find new ways, unconventional and creative solutions [23].

Communication refers to a person's ability to express ideas using words and language that can be accepted both in written and spoken form, such as articulating, explaining, describing, clarifying, listening, questioning, sharing, which are from a learning process [2]. Communication skills are the main factor to be able to understand learning. Teachers, students, and everyone in the educational environment can support each other's learning goals by communicating. In addition, communication has a significant contribution to student success in school aligns with millennial habits and lifestyles that are identical to communication technology if used wisely by students [24].

One of the learning objectives is to master the concept; students must also have good communication skills to convey their knowledge and find. However, based on the research results, many students still do not have good communication skills [5]. Therefore, a method is needed in the form of an appropriate learning approach to develop students' communication skills, and their understanding of concepts can be appropriately conveyed.

III. Material and Method

A. Research Design

This research is systematic literature review. The systematic review is a critical and transparent method for finding, determining, selecting, and synthesizing sources of information from published empirical evidence so that it can answer research questions to be carried out. The systematic review is a research method that based on evidence that has had high credibility in many research disciplines in recent years, including education.

B. Sample and Data Collection

The data used in this research is secondary data. Secondary data applies the documentation method from existing data, not data obtained from direct observation. This research's secondary sources are books, proceedings, and scientific articles in reputable journals accessed on databases in Table I. The keywords used in the search for articles in the SCOPUS database were mobile learning, higher-order thinking skills, and communication skills. The article search was limited from 2012 to 2021, with 30 articles.

TABLE I. THE NUMBER OF PAPERS FROM EACH JOURNAL ANALYZED

Publisher	Frequency	Percentage (%)
https://scholar.google.co.id	3	10
https://doaj.org	5	16.67
https://www.elsevier.com	7	23.33
https://www.tandfonline.com	7	23.33
https://onlinelibrary.wiley.com	4	13.33
https://iopscience.iop.org	4	13.33

C. Analyzing of Data

The main focus of the study in this research is learning science with mobile applications/technology to improve higher-order thinking and communication skills. Therefore, a systematic review requires an appropriate search for the most relevant primary empirical studies. In addition, it is essential to provide details on how the review process was carried out so that the review is transparent and adequate. Therefore, in the following sections, the process of searching, selecting, extracting, and analyzing data is briefly discussed.

- Define research questions to extract the main search.
- Identify relevant keywords in the primary literature search that will use.
- Select various online databases, journals, and conference proceedings to search.
- Conduct a review by reading the article's abstract about the research topic. The author takes descriptive information such as author, year of publication, topic, type of research, and findings.
- Manage results (citations and abstracts) according to the purpose and formulation of the research problem
- Write the results of the review according to the research topic.

Based on our primary research question, we have four main search terms: Mobile Learning, Science, Higher-order thinking skills, and communication skills. Based on the results obtained in an online database search, 30 articles were synthesized and analyzed to answer research questions.

IV. RESULT

A. Mobile Device use in Learning

Mobile devices that should be used in learning are mobile devices that are easy to carry and have easily accessible on and off buttons. Portable devices, such as cell phones and tablets meet this requirement, with the exception of laptops due to their lack of speed in their ease of use. The results of the study report that in learning mobile phones are the most frequently used mobile devices according to the Table II. However, some studies did not identify the device used [25].

TABLE II. MOBILE DEVICES USED IN LEARNING

Mobile Device	Percentage (%)
Phone	34
PDAs	22
Tablets	16
iPads	11
e-books	2
iPods	1

B. Learning Approach in Science to Improve Higher-Order Thinking Skills and Communication Skills

There are 83% of studies that aim to measure student learning improvements who report achieving their stated goals [1]. Exploratory science learning can be carried out outside formal classroom settings, such as in a natural environment [26]. In the previous research review, of all subjects, natural science was the subject whose mobile learning applications were most frequently used in research [27]. Environmental science is the subject matter in science learning that focuses most of the research, followed by geography and physics [1].

Mobile learning positively impacts students' understanding of learning materials and inferring them. Mobile learning should reflect on every learning encounter with communication and collaboration [28]. Furthermore, mobile devices in learning can expand teacher pedagogy and develop students' critical thinking skills and 21st-century skills [29].

However, cultivating communication and high-level skills takes longer, and measuring these abilities requires the right tools and approaches. Below Table III are the learning approach strategies used in mobile learning to improve higher-order thinking skills and communication skills of students in science.

TABLE III. LEARNING APPROACH STRATEGIES USED IN MOBILE LEARNING

Learning Approach	Percentage (%)
Collaborative Learning	25
Inquiry Learning	25
Project-Based Learning	19
Problem-Based Learning	14
Game-Based Learning	10
Flipped Classroom Learning	7

V. DISCUSSION

A. Mobile Device use in Learning

Using mobile devices in learning can help students gradually develop their statistical thinking in everyday life. Students have the concept of independent learning with their character so that interaction occurs continuously [30]. Currently, mobile devices in the classroom as mobile learning has a more significant effect than learning that does not use mobile devices [5].

The affordability of mobile devices that can cross space and time in learning can be utilized to build appropriate learning methods and scenarios. Furthermore, the availability of talented resources and interactive comfort between the natural world and the virtual world can answer the problems of mobile devices in learning [30]. From the development of mobile devices in the last ten years, smartphones, tablets, and computers are the most popular new technology devices in learning.

The existence of various kinds of mobile devices in the world of education that continues to develop, such as cellphones, laptops, students can use them positively to create an interactive knowledge space that focuses on students. Teachers have an important role in supervising the use of mobile devices in learning. Teachers must make students learn with the maximum use of mobile devices [31]. The use of mobile devices in learning helps students to access information, find communication styles, and hone higher-order thinking skills. Therefore, the existence of mobile devices in learning is an essential part of learning innovation.

Previous research by [32] also found that mobile phones were the most frequently used learning devices. In many countries, mobile phone ownership data is more than 100%, with almost everyone owning more than one mobile device; this shows that mobile devices in this study are very affordable [11]. This data also reveals that many devices are used in mobile learning, not depending on what type of device is used but on the access to have the mobile device.

A study conducted by [3] also got the same result: cell phones are the most popular mobile devices for learning. This is because mobile phones have the advantage of multi-tasking; they have various exciting features to help learn, such as photography and video recording. In addition, other features such as GPS, Bluetooth, SMS, Multimedia Messaging Service (MMS), and all kinds of educational software, including the internet and e-books, can be easily accessed by students.

B. *Learning Approach in Science to Improve Higher-Order Thinking Skills and Communication Skills*

1) Collaborative learning

One excellent science learning approach to achieve specific mobile learning targets is collaboration. This is because collaborative learning has the right strategies and methods. Students follow the learning process constantly and continuously to understand the material in each learning process [5], [17]. Collaborative learning is a teaching method where students learn together in a group, and students can help each other for learning purposes. This makes collaborative learning that can provide social experiences and self-development [5].

Collaborative learning is a learning approach to encourage and facilitate students, teachers, classmates, and the community, to be able to interact both inside and outside the classroom. Collaborative learning that occurs outside the classroom can be the primary learning approach. The provision of unstructured learning assignments, learner-centered teaching methods, and mobile learning tools help increase social interaction among peers [30]. In addition, the mobile collaborative learning approach pays excellent attention to student communication in every class activity.

Collaborative learning with mobile technology was the most relevant learning to promote, facilitate, and enhance interaction and collaboration between learners [2]. The collaborative learning approach improves the ability of students to work together in groups, share goals, understand, and discuss to achieve agreed goals [33].

The collaborative learning approach in mobile learning is based on students' ability to engage in productive learning from the information held in study groups that enable students to become researchers and discoverers of knowledge [5]. In many studies, students usually participate in more than one learning activity. For example, a student uses a mobile device to search for new information from his knowledge and share that knowledge with other students. In that case, this builds the learner's constructivist abilities in collaborative learning [3], [25]. This approach will make students ready for digital era learning and generation.

When students study in a collaborative learning environment, students can discuss and exchange information to find answers to problems in their daily lives. Learners will learn from real-life facts and understand current issues regarding the subject matter in learning. Thus students can be involved in various social interactions such as interpersonal, intra-group, intergroup interactions, between humans and the actual and historical environment, and self-reflection [2].

In collaborative learning, students in groups achieve learning goals and learn according to the division of tasks given with full responsibility. The teacher motivates students to be more active in learning activities [1]. When the collaborative learning approach is applied correctly, discussions between students in groups will be smoother and wiser so that students can understand various aspects of specific knowledge and theories in depth. Therefore, students' higher-order thinking skills and communication skills can develop.

Developing knowledge, higher-order thinking, and communication skills should engage students in an open learning environment. Students can learn flexibly and interactively with various alternatives for building knowledge socially in shared learning and investigation groups [30].

2) Inquiry learning

One of the excellent science learning approaches in mobile learning is the inquiry learning approach [5]. Inquiry learning facilitates students to ask questions, conduct investigations or searches, and do experiments to research independently to get the knowledge they need, supported by the theories above [34].

Mobile learning with an inquiry approach is reported to have a diverse focus of study, produces contrasting results, and places greater emphasis on progressive, trustworthy, genuine, and social characteristics [1]. Inquiry learning generally has stages, namely, questioning, inquiry, critical thinking, and problem-solving in which evidence is gathered, findings are reported, explanations are obtained, and conclusions are agreed [35].

Each stage of the inquiry learning approach requires students to think critically and analytically in solving problems. Inquiry-based learning is associated with phenomena by the knowledge that students have previously to help students build new knowledge [36]. It follows the characteristics of students, namely learning through direct experience, primarily through their activities [37]. Therefore, the teacher plays an essential role as a facilitator who provides material and problems to be investigated and guides students in solving problems to understand the existing concepts.

The use of mobile devices in inquiry learning is more effective than learning with lecture methods, independent learning, and cooperative learning [5]. Moreover, integrating mobile devices with inquiry learning can expand students' learning opportunities, the depth of the material understanding and improve students' higher-order thinking skills in completing investigations on scientific problems by synchronizing the results obtained with the explanations given by the teacher [17].

3) Project-based learning

The following science learning approach applied in mobile learning is project-based learning [5]. Project-based learning makes students investigate complex problems and solve them [38].

Project-based learning requires good cooperation and coordination between appropriate group members and appropriate interactions to improve student learning outcomes [39]. Project-based learning is a learning strategy that allows students to benefit from completing learning investigations in the form of thoroughness and accuracy, students have the opportunity to create from their understanding, students can learn and assess learning.

Students are expected to learn and have the ability to design projects in order to solve real-life problems. In addition, project-based learning can improve students' higher-order thinking and communication skills because learning consists of comprehensive, follow-up, and communicative activities to find and combine concepts [5].

4) Problem-based learning

The following science learning approach applied in mobile learning is problem-based learning [5]. Problem-based learning is a learning model that uses a phenomenon or a problem as the primary focus to develop problem-solving and self-organization skills. Problems often used in this learning model are problems in everyday life connected to the subject matter. Problem-based learning requires students to explore knowledge, create ideas, gather information, and solve and answer problems with appropriate solutions [40]. In problem-solving, there is an exchange of knowledge possessed and information collected between each student to solve the problem. The teacher acts as a facilitator to direct the problem so that student discussions focus on solutions.

In the process, the problem-based learning model has beneficial aspects regarding improving students' communication skills. First, students are involved in solving problems. Students' verbal communication skills will develop when they investigate problems, debate theories related to problems, identify problems, propose solutions to problems, analyze and evaluate the problem-solving process, and share experiences related to the process. Second, at each stage of problem-based learning, students' motivation and confidence in communicating can increase. Third, the knowledge obtained by students at the end of the learning process is their original, from the knowledge construction and exploration process they do in problem-based learning. [2].

Problem-based learning encourages students to make meaning from real-life problems that involve higher-order thinking and communication skills. In addition, it involves problem-solving skills, interdisciplinary learning processes, independent learning, and cooperative learning so that all skills involved can be increased [41].

5) Game-based learning

A game-based learning approach can be applied in mobile learning [1]. Games today are developed as entertainment, but some are also developed for learning purposes. Games that are specially designed with technology in education will stimulate the curiosity level of students to learn the existing material [42]. This game-based learning approach can be a choice for teachers to develop interactive learning media. The games aim to support the learning process and increase knowledge in carrying out the learning process compared to the general concept of learning, namely reading.

Game-based learning is a form of learning packaged with games based on specific plans, programs, tools, and equipment prepared by the teacher, and then students are trained in playing the game to achieve the learning objectives set. The game-based learning process is the right tool to represent, visualize, manipulate, and interact with learning content through technology integration [43]. Game-based learning that is packaged excitingly, according to learning objectives, explores and builds concepts, makes students motivated to develop higher-order thinking skills, especially in analysis. In addition, students' communication and collaboration skills are needed in-game activities, especially in groups.

6) Flipped classroom learning

The following science learning approach applied in mobile learning is the flipped classroom [44]. The reverse class is mixed learning, namely learning in two ways; through face-to-face and through virtual or online interactions that combine synchronous learning and asynchronous self-learning. Synchronous learning occurs directly in the classroom, while asynchronous learning is independent learning. The teaching and learning process in the flipped classroom is different; namely, students learn the subject matter that has been given by the teacher via mobile devices at home before the face-to-face class starts.

In learning activities in class, students are asked to do assignments according to the subject matter that has been given by the teacher before, and students discuss subject matter that is not understood with friends and the teacher. By doing assignments at school, when students experience difficulties, they can be directly consulted with their friends or the teacher to solve the problem immediately. A flipped classroom is a learning approach strategy that can minimize the amount of direct instruction in teaching practice but still maximize the interaction process both inside and outside the classroom. This approach strategy makes maximum use of cellular technology [44].

As long as students learn outside the classroom in independent learning, students use mobile devices, such as cellphones, laptops, and computers. This mobile device is beneficial for students in following the learning and is very dependent on it. The flipped classroom approach in its application sharpens students' communication skills [5]. Every teacher and student meets online with the intermediary of a mobile device so that learning runs optimally, two-way communication continues.

The highest percentage of learning approach that used to improve students' higher-order thinking skills and communication skills in science learning are collaborative learning and inquiry learning.

This study has several limitations that reduce the generalizability of the findings: First, this study examines what mobile learning devices are most often used in science learning regardless of the students' economic background. Second, and most importantly, various factors other than mobile devices and learning approaches can improve students' higher-order thinking and communication skills

VI. Conclusion

This research can add to the basis of scientific research and be useful for future researchers because it provides an up-to-date review of mobile learning in science. The most appropriate mobile device used to achieve the learning objectives is a mobile phone, followed by PDAs, tablets, iPad, laptops, e-books, and iPod. Furthermore, it is hoped that this research can be an illustration for other researchers to create innovative learning approaches. The innovative learning approach used in science learning to improve higher-order thinking skills and communication skills is collaborative, inquiry, project-based, problem-based, game-based, and flipped classroom learning. In addition, mobile learning has provided a context in which there are many ways to achieve educational goals, such as independent learning, learning anywhere and anytime, learning to interests, and learning to recognize the characteristics of the student.

VII. Recommendation

Based on the learning approach and theory, the problem faced as a teacher and developer of a learning environment is to ensure that learning will occur adequately and comfortably so that students feel interested in learning. Therefore, one of the essential requirements to fulfill such mobile learning is the learning approach and theory.

Some research that can be done next: First, research on how mobile learning in social learning or comparing the two. Secondly, further research on the most appropriate learning media for mobile learning. And thirdly, research on the effectiveness of implementing approach strategies in mobile learning.

Acknowledgment

The authors thank to Ministry of Education and Culture, Research and Technology, Republic of Indonesia for financial support during this research PMDSU Scholarship program.

References

[1] M. Bano, D. Zowghi, M. Kearney, S. Schuck, and P. Aubusson, “Mobile learning for science and mathematics school education: A systematic review of empirical evidence,” Comput. Educ., vol. 121, pp. 30–58, 2018, doi: 10.1016/j.compedu.2018.02.006.

[2] J. Mou and J. F. Cohen, “A longitudinal study of trust and perceived usefulness in consumer acceptance of an e-service: The case of online health services,” in Proceedings of the 18 th Pacific Asia Conference on Information Systems, 2014, p. 258.

[3] H. Hamidi and A. Chavoshi, “Analysis of the essential factors for the adoption of mobile learning in higher education: A case study of students of the University of Technology,” Telemat. Informatics, vol. 35, no. 4, pp. 1053–1070, 2018, doi: 10.1016/j.tele.2017.09.016.

[4] J. Gikas and M. M. Grant, “Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media,” Internet High. Educ., vol. 19, pp. 18–26, 2013.

[5] M. Shao and X. Liu, “Impact of the Flipped Classroom on Students’ Learning Performance via Meta-Analysis,” Open J. Soc. Sci., vol. 09, no. 09, pp. 82–109, 2021, doi: 10.4236/jss.2021.99007.

[6] J. Cheon, S. Lee, S. M. Crooks, and J. Song, “An investigation of mobile learning readiness in higher education based on the theory of planned behavior,” Comput. Educ., vol. 59, no. 3, pp. 1054–1064, 2012, doi: 10.1016/j.compedu.2012.04.015.

[7] H. Crompton, B. Olszewski, and T. Bielefeldt, “The mobile learning training needs of educators in technology-enabled environments,” Prof. Dev. Educ., vol. 42, no. 3, pp. 482–501, 2016, doi: 10.1080/19415257.2014.1001033.

[8] P. A. Ertmer and A. T. Ottenbreit-Leftwich, “Teacher technology change: How knowledge, confidence, beliefs, and culture intersect,” J. Res. Technol. Educ., vol. 42, no. 3, pp. 255–284, 2010, doi: 10.1080/15391523.2010.10782551.

[9] M. Milrad, L. Wong, M. Sharples, G. J. Hwang, G. Looi, and H. Ogata, Seamless learning: an international perspective on next-generation technology-enhanced learning. 2013.

[10] N. Selwyn, “Looking beyond learning: Notes towards the critical study of educational technology,” J. Comput. Assist. Learn., vol. 26, no. 1, pp. 65–73, 2010, doi: 10.1111/j.1365-2729.2009.00338.x.

[11] A. Tsinakos and M. Ally, Global Mobile Learning Implementations and Trends, Open Book. China: TV University Press, 2013.

[12] C. Yu, S. J. Lee, and C. Ewing, “Mobile Learning : Emerging Trends , Issues , and Challenges in Teaching and Learning,” no. 2003, pp. 2126–2136, 2011.

[13] H. Jenkins and M. Ito, Participatory culture in a networked era. A conversation on youth, learning, commerce and politics. 2016.

[14] J. H. Kuznekoff, S. Munz, and S. Titsworth, “Mobile phones in the classroom: Examining the effects of texting, twitter, and message content on student learning,” Commun. Educ., vol. 64, no. 3, pp. 344–365, 2015.

[15] R. Arum and J. Roksa, Academically adrift: Limited learning on college campuses. 2011.

[16] E. Baran, “A Review of Research on Mobile Learning in Teacher Education,” Educ. Technol. Soc., vol. 17, no. 4, pp. 17–32, 2014.

[17] L. F. M. G. Pedro, C. M. M. de O. Barbosa, and C. M. das N. Santos, “A critical review of mobile learning integration in formal educational contexts,” Int. J. Educ. Technol. High. Educ., vol. 15, no. 1, 2018, doi: 10.1186/s41239-018-0091-4.

[18] S. Hao, V. P. Dennen, and L. Mei, “Influential factors for mobile learning acceptance among Chinese users,” Educ. Technol. Res. Dev., vol. 65, no. 1, pp. 101–123, 2017, doi: 10.1007/s11423-016-9465-2.

[19] N. Mallat, M. Rossi, V. K. Tuunainen, and A. Oorni, “The impact of use context on mobile services acceptance: The case of mobile ticketing,” Inf. Manag., vol. 46, no. 3, pp. 190–195, 2009.

[20] I. Z. Ichsan, D. V. Sigit, M. Miarsyah, A. Ali, W. P. Arif, and T. A. Prayitno, “HOTS-AEP: Higher order thinking skills from elementary to master students in environmental learning,” Eur. J. Educ. Res., vol. 8, no. 4, pp. 935–942, 2019, doi: 10.12973/eu-jer.8.4.935.

[21] L. O. Wilson, “Anderson and Krathwohl Bloom’s Taxonomy Revised Understanding the New Version of Bloom’s Taxonomy,” Second Princ., pp. 1–8, 2016.

[22] I. Wayan Widana. “Higher Order Thinking Skills Assessment (Hots).” Pgri Bali, vol. 3, no. 1, pp. 32–44, 2017.

[23] S. M. Brookhart, How to assess higher-order thinking skills in your classroom. 2010.

[24] A. Khoiri et al., “4Cs Analysis of 21st Century Skills-Based School Areas,” J. Phys. Conf. Ser., vol. 1764, no. 1, 2021, doi: 10.1088/1742-6596/1764/1/012142.

[25] H. Crompton, D. Burke, and K. H. Gregory, “The use of mobile learning in PK-12 education: A systematic review,” Comput. Educ., vol. 110, pp. 51–63, 2017, doi: 10.1016/j.compedu.2017.03.013.

[26] J. M. Zydney and Z. Warner, “Mobile apps for science learning: Review of research,” Comput. Educ., vol. 94, pp. 1–17, 2016, doi: 10.1016/j.compedu.2015.11.001.

[27] M. Liu, R. Scordino, R. Geurtz, C. Navarrete, Y. Ko, and M. Lim, “A look at research on mobile learning in K–12 education from 2007 to the present,” J. Res. Technol. Educ., vol. 46, no. 4, pp. 325–372, 2014.

[28] D. Frohberg, C. Göth, and G. Schwabe, “Mobile Learning projects - a critical analysis of the state of the art: Original article,” J. Comput. Assist. Learn., vol. 25, no. 4, pp. 307–331, 2009, doi: 10.1111/j.1365-2729.2009.00315.x.

[29] M. M. Terras and J. Ramsay, “The five central psychological challenges facing effective mobile learning,” Br. J. Educ. Technol., vol. 43, no. 5, pp. 820–832, 2012, doi: 10.1111/j.1467-8535.2012.01362.x.

[30] Q. K. Fu and G. J. Hwang, “Trends in mobile technology-supported collaborative learning: A systematic review of journal publications from 2007 to 2016,” Comput. Educ., vol. 119, pp. 129–143, 2018, doi: 10.1016/j.compedu.2018.01.004.

[31] M. S. K. Batiibwe and F. E. K. Bakkabulindi, “Technological Pedagogical Content Knowledge (Tpack) as a Theory on Factors of the Use of ICT in Pedagogy: A Review of Literature,” in South Africa International Conference on Education, 2016, pp. 228–241.

[32] W. H. Wu, Y. C. Jim Wu, C. Y. Chen, H. Y. Kao, C. H. Lin, and S. H. Huang, “Review of trends from mobile learning studies: A meta-analysis,” Comput. Educ., vol. 59, no. 2, pp. 817–827, 2012, doi: 10.1016/j.compedu.2012.03.016.

[33] M. Martín del Pozo, V. Basilotta Gómez-Pablos, and A. García-Valcárcel Muñoz-Repiso, “A quantitative approach to pre-service primary school teachers’ attitudes towards collaborative learning with video games: previous experience with video games can make the difference,” Int. J. Educ. Technol. High. Educ., vol. 14, no. 1, 2017, doi: 10.1186/s41239-017-0050-5.

[34] K. Burden and M. Kearney, “Future scenarios for mobile science learning,” Res. Sci. Educ., vol. 46, no. 2, pp. 287–308, 2016.

[35] J. C. Marshall, R. Horton, B. L. Igo, and D. M. Switzer, “K-12 science and mathematics teachers’ beliefs about and use of inquiry in the classroom.,” Int. J. Sci. Math. Educ., vol. 7, pp. 575–596, 2009.

[36] R. Utami and E. Rohaeti, “Students’ Generic Science Skills in Chemistry Learning Using Inquiry-Based Learning,” vol. 317, no. IConProCS, pp. 234–238, 2019, doi: 10.2991/iconprocs-19.2019.49.

[37] M. Nazar, R. F. I. Rahmayani, and Z. Yulia, “the Development of Students’ Worksheet Based on Guided Inquiry in Corrosion Matter,” Edusains, vol. 10, no. 2, pp. 287–294, 2018, doi: 10.15408/es.v10i2.8699.

[38] E. C. Miller, S. Severance, and J. Krajcikc, “Motivating teaching, sustaining change in practice: design principles for teacher learning in project-based learning contexts,” J. Sci. Teacher Educ., vol. 32, no. 7, pp. 757–779, 2021.

[39] K. Juuti, J. Lavonen, V. Salonen, K. Salmela-Aro, B. Schneider, and J. Krajcik, “A Teacher–Researcher Partnership for Professional Learning: Co-Designing Project-Based Learning Units to Increase Student Engagement in Science Classes,” J. Sci. Teacher Educ., vol. 32, no. 6, pp. 625–641, 2021, doi: 10.1080/1046560X.2021.1872207.

[40] R. D. Anazifa and Djukri, “Project- based learning and problem- based learning: Are they effective to improve student’s thinking skills?,” J. Pendidik. IPA Indones., vol. 6, no. 2, pp. 346–355, 2017, doi: 10.15294/jpii.v6i2.11100.

[41] C. Tosun and Y. Taskesenligil, “The effect of problem-based learning on undergraduate students’ learning about solutions and their physical properties and scientific processing skills,” Chem. Educ. Res. Pract., vol. 14, no. 1, pp. 36–50, 2013, doi: 10.1039/c2rp20060k.

[42] P. Y. Chen, G. J. Hwang, S. Y. Yeh, Y. T. Chen, T. W. Chen, and C. H. Chien, “Three decades of game-based learning in science and mathematics education: an integrated bibliometric analysis and systematic review,” J. Comput. Educ., 2021.

[43] M. E. Eltahir, N. R. Alsalhi, S. Alqatawneh, H. A. Alqudah, and M. Jaradat, “The impact of game-based learning (GBL) on students’ motivation, engagement and academic performance on an Arabic language grammar course in higher education,” Educ. Inf. Technol., vol. 26, no. 4, pp. 3251–3278, 2021.

[44] S. C. Chang and G. J. Hwang, “Impacts of an augmented reality-based flipped learning guiding approach on students’ scientific project performance and perceptions,” Comput. Educ., vol. 125, pp. 226–239, 2018, doi: 10.1016/j.compedu.2018.06.007.

Superpixel Sizes using Topology Preserved Regular Superpixel Algorithm and their Impacts on Interactive Segmentation

Kok Luong Goh[1], Soo See Chai[2], Giap Weng Ng[3], Muzaffar Hamzah[4]
Faculty of Science and Technology, i-CATS University College, Kuching, Malaysia[1]
Faculty of Computer Science and Information Technology, University Malaysia Sarawak, Kuching, Malaysia[2]
Faculty of Computing and Informatics, University Malaysia Sabah, Kota Kinabalu, Malaysia[1, 3, 4]

***Abstract*—Interactive Image Segmentation is a type of semi-automated segmentation that uses user input to extract the object of interest. It is possible to speed up and improve the end result of segmentation by using pre-processing steps. The use of superpixels is an example of a pre-processing step. A superpixel is a collection of pixels with similar properties such as texture and colour. Previous research was conducted to assess the impact of the number of superpixels (based on SEEDS superpixel aglorithms) required to achieve the best segmentation results. The study, however, only examined one type of input (strokes) and a small number of images. As a result, the goal of this study is to extend previous work by performing interactive segmentation with input strokes and a combination of bounding box and strokes on images from Grabcut image data sets generated by Topology preserved regular superpixel (TPRS). Based on our findings, an image with 1000 to 2500 superpixels and a combination of bounding box and strokes will help the interactive segmentation algorithm produce a good segmentation result. Finally, the size of the superpixels would influence the final segmentation results as well as the input type.**

Keywords—Image segmentation; superpixel; input type; interactive segmentation

I. Introduction

Computer vision is an artificial intelligence subfield that teaches computers to understand and interpret their visual environment. Image processing is a subset of computer vision that entails enhancing or extracting useful information from images. Image segmentation, a subprocess of image processing, on the other hand, is a technique that allows humans to extract objects of interest from images. Image segmentation can be done manually, semi-automatically, or fully automatically.

In automated segmentation, there is no user involvement. Semi-automated segmentation, also known as interactive segmentation, requires little to no user intervention during the segmentation process. The ultimate goal of image segmentation is to fully automate the process. However, because of the image complexity, automated segmentation continues to face significant challenges in producing satisfactory results. As a result, for better image segmentation results, semi-automated or interactive image segmentation methods are preferred.

Traditionally, image segmentation is performed using information from each individual pixel in the image. This process, however, consumes a significant amount of processing power. Ren and Malik [1] proposed superpixel as a solution to this issue. A superpixel is a group of pixels with similar properties, such as texture or colour. The introduction of the superpixel altered the segmentation processing steps.

Since the introduction of superpixels, many interaction segmentation methods have used them as part of the pre-processing phase [2-5]. However, different interactive segmentation algorithms employ different superpixel algorithms of varying sizes. For instance, [6] and [7] used the SLIC superpixel algorithm [8] with 1000 and 2000 superpixels superpixel per image, respectively. In addition, [9] used a meant-shift superpixel algorithm with 100 superpixels per image. As can be seen, no standard method exists in interactive segmentation for determining the size of a superpixel. As a result, a study [10] was conducted using MSRM interactive segmentation [11] and the SEEDS superpixel algorithm[12] to assess the effect of the number of superpixels required to achieve the best segmentation results. According to the study, 500 superpixels per image were the optimal size for achieving a good segmentation result. The study, however, only looked at one type of input (strokes) and a small number of images. Therefore, the purpose of this paper is to expand on the previous study on the following:

- Using all images from the Grabcut dataset [13] (50 images).
- Using different superpixel algorithm e.g., Topology preserved regular superpixel (TPRS) [14].
- Using various input types, such as bounding box and strokes.

The study's findings will be compared to state-of-the-art interactive algorithms on a variety of metrics.

The study's findings will provide information on the optimal size of the superpixel for TPRS based on the input types utilised by the MSRM interactive segmentation algorithms. In addition, it will serve as an extra guideline when TPRS superpixels are used with other segmentation algorithms.

In the following section, interactive segmentation will be discussed, followed by an overview of the various user input types used in interactive segmentation, maximal similarity-based region merging, and topology preserved regular superpixel algorithms.

II. Literature Review

Interactive image segmentation has been used in a variety of applications. For example, a tool for medical volume images has been created (SmartPaint [15] and MRI for orthopaedic surgery[16]. In the field of remote sensing, a segmentation tool for lithological boundary detection has also been developed [17]. It has also proven to be very useful in agriculture, assisting farmers in detecting crop diseases [18].

As stated in the preceding paragraph, the user will provide guidance in order for the segmentation system to extract the object of interest in interactive segmentation. The general process of interactive segmentation is summarized below:

Step 1: The user will provide information about the context and the object of interest.

Step 2: The segmentation system will generate a segmentation result based on the user's input.

Step 3: The user will evaluate the results, and if the user is satisfied, the process will be completed. Otherwise, the user will provide more background and object of interest information until the system produces a satisfactory segmented result.

Existing work [19] distinguishes between interactive and semi-automatic segmentation by involving the user in both the initialization and post-processing stages of the segmentation process iteratively, whereas semi-automatic segmentation only involves the user in the initialization stage. This study combines the two terms and defines interactive segmentation as any segmentation that requires user input. This study, on the other hand, will concentrate on the involvement of user input during the initialization stage.

Various input types are used in interactive segmentation to provide information about the background and object of interest. Some examples of these input formats are as follows:

- Strokes [20-24]: The user must apply stroke(s) to the image's object of interest and background.
- Seed point [25-29]: The seed points must be placed on the image's background and object of interest by the user.
- Bounding box [13, 30-34]: The user must position the bounding box around the object of interest within the image.

The following section explains Maximal Similarity-based Region Merging (MSRM) and Topology Preserved regular superpixels (TPRS).

A. Maximal Similarity-based Region Merging (MSRM)

Maximal Similarity-based Region Merging algorithm [11] is based on region merging. The image is first converted into superpixels using mean shift segmentation. The contour of the object is then extracted based on the labelling of non-marked regions as region of interest or background. Fig. 1(a) shows the superpixels of the image with strokes on the background and object of interest, and (b) shows the segmentation result.

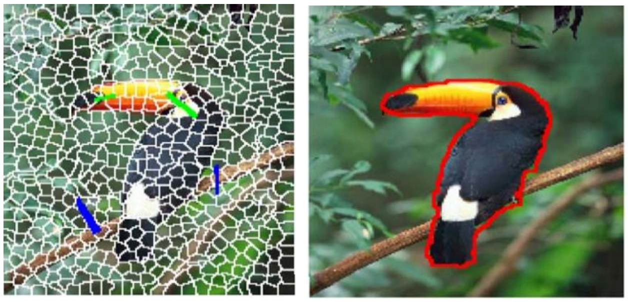


Fig. 1. The Algorithm's Segmentation Process [11]: (a) Superpixel Strokes Entered by users (b): The Segmentation Outcome.

B. Topology Preserved Regular Superpixel (TPRS)

Topology preserved regular superpixel (TPRS) [14] is a path-based method that partitions an image into superpixels by connecting seed points via pixel path. It begins by arranging initial seeds on a lattice grid and associating them with appropriate pixels on the boundary map. It then relocates each seed to the pixel with the highest locally maximal edge magnitudes, taking into account both the distance term and the probability term. Finally, it finds the local optimal path between vertical and horizontal seed pairs.

III. Experimental Setting

To determine the best superpixel size, TPRS will generate a test image with five superpixel sizes: 500, 1000, 1500, 2000, and 2500. Aside from that, each superpixel image will be paired with the following user input type:

- s1: 1 foreground stroke and 3 background strokes.
- s2: multiple foreground and background strokes.
- m2: 1 background bounding box and 2 foreground strokes.
- m3: 1 background bounding box and 2 foreground strokes.

The user input type and superpixel image as well as test image will be fed into the MSRM interactive segmentation algorithm in order to generate segmentation results. Fig. 2 shows some of test image with ground truth (see (a) and (b)). Aside from that, s1 and s2 input strokes from [35] were chosen, as shown in Fig. 2(c) and (d). Fig. 2(e) and (f) show the use of a bounding box with two and three foreground strokes, respectively. Fig. 3 shows the superpixel image generated by TPRS with sizes of 500 in (a) and 2500 in (b).

To assess the efficacy of superpixel algorithms in interactive segmentation, the error rate, F-score, and Jaccard indexes used by [31, 36-38]. Error rate (ERR) is the percentage of pixels placed in an incorrect region which is shown as below equation:

$$ERR = 1 - \left(\frac{TP+TN}{TP+TN+TP+FN}\right) \quad (1)$$

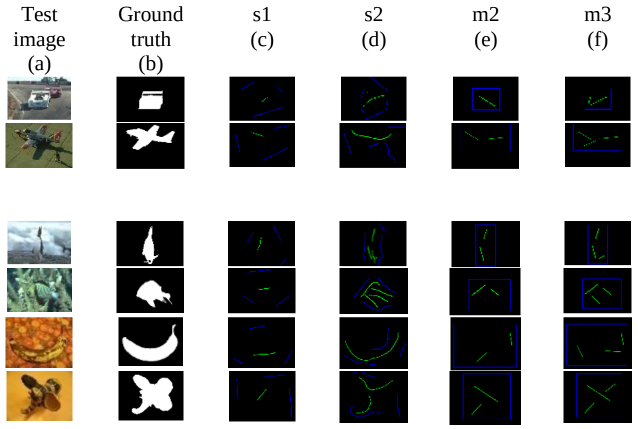


Fig. 2. (a) Test Image. (b) Ground Truth. (c) Simple Stroke. (d) Complex Stroke. (e) Bounding Box with Two Foreground Strokes. (f) Bounding Box with Three Foreground Strokes.

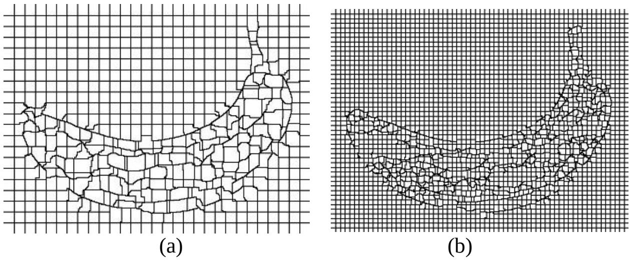


Fig. 3. Superpixel Image of Banana1 from the Dataset which had been Generated by TPRS. (a) 500 (b) 2500.

However, error rate takes into account the percentage of pixels that accurately map to the background information. As a result, the F-score and Jaccard Index are included.

F-score is equivalent to Dice Coefficient. The F-score is also known as the F1-Score or F-Measure. It is equal to 2 * the Area of Overlap divided by the total number of pixels in both images.

$$P = \frac{TP}{TP+FP} \qquad (2)$$

$$R = \frac{TP}{TP+FN} \qquad (3)$$

$$F = 2 * \left(\frac{P*R}{P+R}\right) \qquad (4)$$

The Jaccard index, also known as the Intersection over Union (IoU) metric. It is the ratio of the number of pixels that are shared by X and Y to the total number of pixels in X and Y. In this case, X and Y represent the segmented image and ground truth, respectively. The Jaccard index/ IoU formulation is depicted as follows:

$$J/IOU = \frac{TP}{TP+FP+FN} \qquad (5)$$

IV. Results and Discussion

To have better understanding of the result generated from various size as well as input type, the table below (see Table I) had divided the result based on the superpixel size, s and user input type, t. following by Error rate, e, f-score, f and Jaccard index, j. The error rate, F-score, and Jaccard index are also depicted graphically in Fig. 4 and 5.

TABLE I. Overall Image Segmentation Result based on Three Metrics: Error Rate, e; f, f-score; and Jaccard Index, j Performed by MSRM by using TPRS Superpixel Algorithms that Generated Different Numbers of Superpixels, s Ranging between 500, 1000, 1500, 2000, and 2500 by using Different Types of Input Types: m2 (Bounding Box with Two Foreground Strokes), m3 (Bounding Box with Three Foreground Strokes), s1 (Simple Strokes) and s2 (Complex Strokes)

s	t	e ↓	p ↑	r ↑	f ↑	j ↑
500	m2	0.037	0.943	0.842	0.882	0.803
500	m3	0.030	0.939	0.893	0.913	0.845
1000	m2	0.034	0.947	0.856	0.890	0.816
1000	m3	0.025	0.943	0.919	0.929	0.872
1500	m2	0.036	0.955	0.833	0.878	0.802
1500	m3	0.027	0.954	0.886	0.915	0.851
2000	m2	0.030	0.950	0.867	0.896	0.830
2000	m3	0.025	0.950	0.904	0.923	0.863
2500	m2	0.032	0.954	0.842	0.881	0.811
2500	m3	0.024	0.952	0.898	0.921	0.863
500	s1	0.059	0.911	0.776	0.805	0.712
500	s2	0.036	0.907	0.901	0.900	0.825
1000	s1	0.064	0.935	0.745	0.789	0.695
1000	s2	0.029	0.941	0.914	**0.924**	**0.864**
1500	s1	0.066	0.928	0.740	0.778	0.685
1500	s2	0.030	0.949	0.901	0.921	0.859
2000	s1	0.068	0.953	0.704	0.759	0.668
2000	s2	0.029	0.943	0.902	0.918	0.856
2500	s1	0.064	0.943	0.699	0.751	0.658
2500	s2	**0.028**	0.958	0.891	0.919	0.857

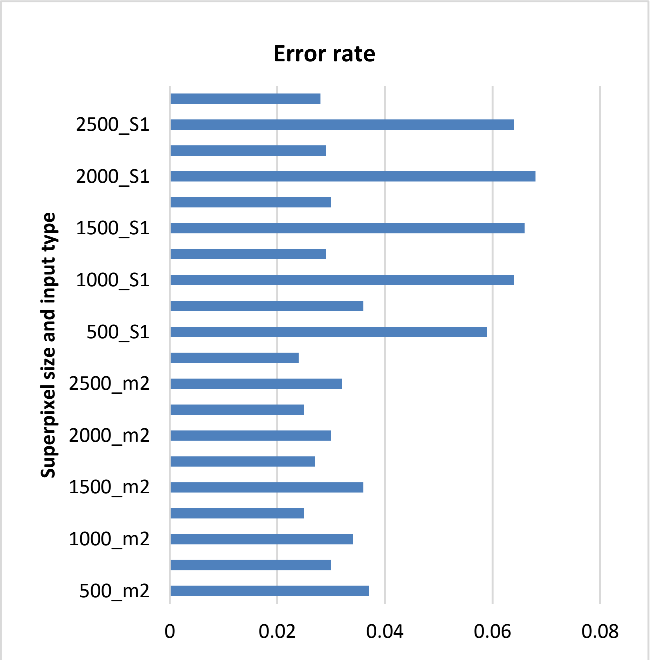


Fig. 4. Segmentation Results based on Error Rate using Various user Input T and Superpixel Size.

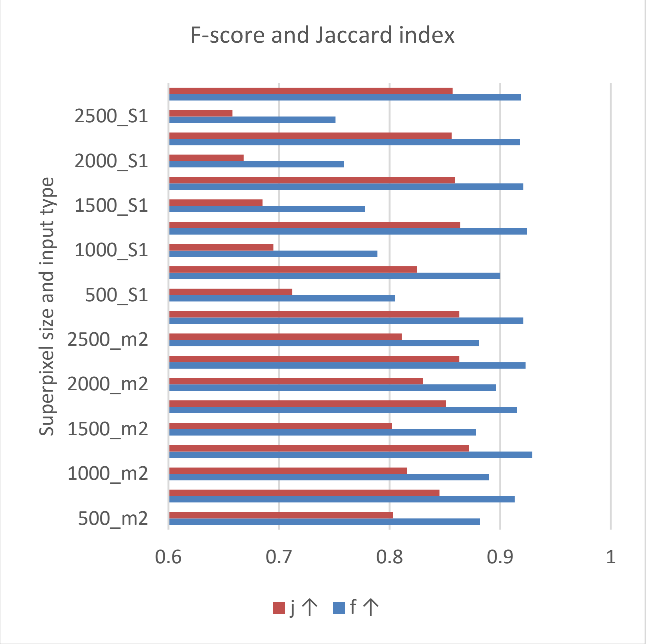


Fig. 5. Segmentation Results based on F-score and Jaccard Index using Various user Input Types and Superpixel Size.

From the Table I, it can summarize the finding of research study as below:

- In term of input type, bounding box input type had performed better than strokes with error rate 0.024 and 0.028, f-score 0.929 and 0.924, and Jaccard index 0.872, 0.864, respectively.
- Regardless of input type, superpixel size of 1000 had outperformed than other superpixel size in term of Jaccard index and F-score. On the other hand, if using error rate as a metric, size of 2500 is much better than size of 1000. However, the difference between only 0.001. Also, if comparing using f-score (0.005-0.008) and Jaccard index (0.007-0.009). It can conclude that size of 1000 should be an optimum size for TPRS.
- The increasing of input strokes for foreground had improved the segmentation results regardless of bounding box or strokes.
- The difference between the highest and lowest result for each category (s1, s2, m2, m3) of each matric are 0.006 and 0.009, error rate, 0.016 and 0.054, f-score and 0.024 and 0.054, Jaccard index.
- Overall, the difference between the highest and lowest result of all are 0.044, error rate, 0.178, f-score, and 0.214, Jaccard index.

The individual image results based on the 1000 and 2500 superpixels for strokes and bounding box categories shown in Table II. These two sizes from two categories were chosen due to the fact that they had achieved the best result among others. From the Table II, it is shown that 153077 had achieved high error rate (>0.1) if using bounding box while 189080 and stone2 if using strokes.

TABLE II. INDIVIDUAL IMAGE RESULT (ERROR RATE, E) BASED ON THE 1000 AND 2500 SUPERPIXELS FOR STROKES AND BOUNDING BOX CATEGORIES

filename	1000_m3	2500_m3	1000_s2	2500_s2
106024	0.023	0.019	0.016	0.016
124084	0.030	0.026	0.030	0.025
153077	0.142	0.131	0.058	0.089
153093	0.030	0.043	0.033	0.034
181079	0.030	0.017	0.030	0.054
189080	0.020	0.020	0.135	0.131
208001	0.012	0.009	0.010	0.012
209070	0.045	0.051	0.030	0.041
21077	0.028	0.033	0.016	0.020
227092	0.013	0.024	0.013	0.032
24077	0.030	0.035	0.024	0.047
271008	0.030	0.030	0.015	0.026
304074	0.031	0.058	0.044	0.019
326038	0.031	0.027	0.060	0.054
37073	0.032	0.047	0.020	0.030
376043	0.026	0.029	0.017	0.021
388016	0.017	0.014	0.059	0.018
65019	0.010	0.014	0.006	0.018
69020	0.052	0.038	0.066	0.045
86016	0.006	0.007	0.032	0.007
banana1	0.037	0.031	0.028	0.015
banana2	0.035	0.030	0.018	0.057
banana3	0.025	0.028	0.024	0.029
book	0.029	0.073	0.024	0.026
bool	0.028	0.028	0.013	0.039
bush	0.033	0.020	0.016	0.024
ceramic	0.071	0.035	0.018	0.026
cross	0.013	0.009	0.009	0.020
doll	0.017	0.004	0.003	0.018
elefant	0.040	0.010	0.010	0.015
flower	0.009	0.034	0.049	0.011
fullmoon	0.003	0.018	0.024	0.003
grave	0.016	0.010	0.008	0.009
llama	0.016	0.007	0.038	0.064
memorial	0.020	0.009	0.016	0.022
music	0.014	0.014	0.011	0.010
person1	0.013	0.012	0.020	0.010
person2	0.009	0.026	0.007	0.011
person3	0.010	0.006	0.015	0.010
person4	0.018	0.015	0.007	0.020
person5	0.008	0.047	0.017	0.018
person6	0.017	0.006	0.005	0.024
person7	0.008	0.007	0.012	0.006
person8	0.016	0.007	0.007	0.017
scissors	0.045	0.034	0.017	0.064
sheep	0.005	0.029	0.030	0.004
stone1	0.009	0.026	0.021	0.007
stone2	0.007	0.037	0.178	0.007
teddy	0.016	0.013	0.049	0.034
tennis	0.033	0.018	0.037	0.030

TABLE III. PERFORMANCE COMPARISON WITH OTHER STATE-OF-ART INTERACTIVE SEGMENTATION ALGORITHMS BASED ON ERROR RATE, E AND F-SCORE, F

Reference algorithms	**e**	**Reference algorithms**	**f**
Extreme points [37]	0.023	GSC(reported in [39])	0.966
Diffusive likelihood [38]	0.023	BNQ [39]	0.947
2500_m3	0.024	Region-based nonparametric [40]	0.942
1000_m3	0.025	RW (reported in [39]	0.937
Probabilistic diffusion [41]	0.027	supercut average [33]	0.9356
1000_m3	0.025	NHO (reported in [40]	0.934
2500_s2	0.028	IGC (reported in [39]	0.933
1000_s2	0.029	Densecut [34]	0.932
Label propagation [42]	0.0321	1000_m3	0.929
Object Extraction from Bounding Box Prior [43]	0.032	LS (reported in [39])	0.926
Xia (reported in [43]	0.033	1000_s2	0.924
SBT with AT(reported in [43]	0.033	onecut(reported in [32]	0.923
RW with AT(reported in [44]	0.033	2500_m3	0.921
NHO (reported in [38]	0.034	SAL [45]	0.9207
MGC (reported in [38]	0.026	GrabCut(GMM) (reported in [34])	0.909
DEEPGC (reported in [37]	0.034	2500s2	0.919
PLL (reported in [38]	0.0349	grabcut (reported in [32]	0.916
TAM (reported in [38]	0.0364	milcut avg [33]	0.9136
Milcut (reported in[37])	0.036	ppbc (reported [32])	0.91
BoxPrior (reported in [37]	0.037	gbmr (reported in [40]	0.91
		RC (reported in [45])	0.9094
		RWR(reported in [45]	0.9057
		LC (reported in [45]	0.8839
		loosecut [32]	0.882
		ABS (reported in [45])	0.7155

TABLE IV. PERFORMANCE COMPARISON WITH OTHER STATE-OF-ART INTERACTIVE SEGMENTATION ALGORITHMS BASED ON JACCARD INDEX, J

Reference algorithms	**j**	**Reference algorithms**	**j**
grabcut (reported in [46])	0.91	SGML [47]	0.81
ppbc (reported in [46])	0.91	PD (reported in [47])	0.8
nc_cut(reported in [46])	0.91	onecut (reported in [46])	0.8
1000_m3	0.872	milcut (reported in [46])	0.8
1000_s2	0.864	RTPG (reported in [48])	0.76
2500_m3	0.863	SD (reported in [48])	0.76
2500_s2	0.857	NRW (reported in [47])	0.72
LPD [48]	0.85	TPG (reported in [47])	0.7
nc_cut0 (reported in [46])	0.85	LC (reported in [47])	0.68
FGML [47]	0.82	SMRW (reported in [47])	0.62

When compared to state-of-the-art interactive segmentation algorithms, 2500 m3 achieved 0.024, which is 0.001 higher than the best reference algorithm in the error rate category, which is a training-based algorithm (see Table III). It did, however, outperforms other training-based algorithms like Label propagation [43] (0.0321), Xia [44] (0.033), SBT with AT [44] (0.033,) RW with AT [45] (0.033), and DEEPGC [38] (0.034). Apart from this, in terms of the F-score (see Table III) and Jaccard index (see Table IV), there is a 4% difference between the best reference algorithm and our algorithm.

V. CONCLUSION AND FUTURE DIRECTION

The TPRS superpixels algorithm was used in this paper to generate various sizes of superpixels images on interactive segmentation algorithms, MSRM using strokes and a combination of bounding box with strokes as user input. The entire test images from the Grabcut dataset were used in the testing. Evaluation matrices such as error rate, e, f-score, f, and Jaccard index, J were used to evaluate the generated results. Overall, the interactive segmentation algorithm was able to achieve an optimal segmentation result by bounding the box with three strokes and using superpixel sizes of 1000 and 2500. This discovery will be useful for researchers who want to use superpixels, particularly on TPRS, with appropriate settings on the number of superpixels in an image on the interactive segmentation algorithm. The segmentation results, on the other hand, were compared to the results of the previous study, which determined that 500 is the optimal number of superpixels to use in interactive segmentation. The difference could be due to the number of test images used as well as the input types. Finally, we make several recommendations for future research:

- Comparative study of various types of superpixel algorithms with diverse input types on different interactive segmentation algorithms.
- Previous studies [49, 50] has found that image complexity can affect segmentation results. Existing interactive image segmentation algorithms have not addressed the relationship between image complexity and input type, as well as superpixel and size, and this can be investigated further.

ACKNOWLEDGMENT

The authors would like to thank Faculty of Computer Science and Information Technology, University of Malaysia Sarawak and Faculty of Science and Technology, i-CATS University College, for providing the facilities for this study.

REFERENCES

[1] Ren, X. and J. Malik. Learning a classification model for segmentation. in Computer Vision, IEEE International Conference on. 2003. IEEE Computer Society.

[2] Zhou, C., et al., An efficient two-stage region merging method for interactive image segmentation. Computers & Electrical Engineering, 2016. 54: p. 220-229.

[3] Ding, J.-J., et al. Real-time interactive image segmentation using improved superpixels. in 2015 IEEE International Conference on Digital Signal Processing (DSP). 2015. IEEE.

[4] Panagiotakis, C., et al., Interactive image segmentation based on synthetic graph coordinates. Pattern Recognition, 2013. 46(11): p. 2940-2952.

[5] Wu, J., et al. Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

[6] Lu, H., et al., Spectral segmentation via midlevel cues integrating geodesic and intensity. IEEE transactions on cybernetics, 2013. 43(6): p. 2170-2178.

[7] Gueziri, H.-E., M.J. McGuffin, and C. Laporte, A generalized graph reduction framework for interactive segmentation of large images. Computer Vision and Image Understanding, 2016. 150: p. 44-57.

[8] Achanta, R., et al., Slic superpixels. 2010.

[9] Jian, M. and C. Jung, Interactive image segmentation using adaptive constraint propagation. IEEE transactions on image processing, 2016. 25(3): p. 1301-1311.

[10] Goh, K.L., et al. Sizes of Superpixels and their Effect on Interactive Segmentation. in 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET). 2021. IEEE.

[11] Ning, J., et al., Interactive image segmentation by maximal similarity based region merging. Pattern Recognition, 2010. 43(2): p. 445-456.

[12] Bergh, M.V.d., et al. Seeds: Superpixels extracted via energy-driven sampling. in European conference on computer vision. 2012. Springer.

[13] Rother, C., V. Kolmogorov, and A. Blake, " GrabCut" interactive foreground extraction using iterated graph cuts. ACM transactions on graphics (TOG), 2004. 23(3): p. 309-314.

[14] Tang, D., H. Fu, and X. Cao. Topology preserved regular superpixel. in 2012 IEEE International Conference on Multimedia and Expo. 2012. IEEE.

[15] Malmberg, F., et al., SmartPaint: a tool for interactive segmentation of medical volume images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2017. 5(1): p. 36-44.

[16] Ozdemir, F., et al., Interactive segmentation in MRI for orthopedic surgery planning: bone tissue. International Journal of Computer Assisted Radiology and Surgery, 2017. 12(6): p. 1031-1039.

[17] Yathunanthan Vasuki, E.-J.H., Peter Kovesi, Steven Micklethwaite, An interactive image segmentation method for lithological boundary detection: A rapid mapping tool for geologists. Computers & Geosciences, 2017. 100: p. 27-40.

[18] Ma, J., et al., A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. Computers and Electronics in Agriculture, 2017. 142: p. 110-117.

[19] Heckel, F., et al., Interactive 3D medical image segmentation with energy-minimizing implicit functions. Computers & Graphics, 2011. 35(2): p. 275-287.

[20] Zhang, J., J. Zheng, and J. Cai. A diffusion approach to seeded image segmentation. in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010. IEEE.

[21] Kim, T.H., K.M. Lee, and S.U. Lee. Generative image segmentation using random walks with restart. in European conference on computer vision. 2008. Springer.

[22] Bai, X. and G. Sapiro. A geodesic framework for fast interactive image and video segmentation and matting. in 2007 IEEE 11th International Conference on Computer Vision. 2007. IEEE.

[23] Wang, T., H. Wang, and L. Fan, A weakly supervised geodesic level set framework for interactive image segmentation. Neurocomputing, 2015. 168: p. 55-64.

[24] Ding, Z., et al., Adaptive fusion with multi-scale features for interactive image segmentation. Applied Intelligence, 2021: p. 1-12.

[25] Adams, R. and L. Bischof, Seeded region growing. IEEE Transactions on pattern analysis and machine intelligence, 1994. 16(6): p. 641-647.

[26] Meena, S., K. Palaniappan, and G. Seetharaman. User driven sparse point-based image segmentation. in 2016 IEEE International Conference on Image Processing (ICIP). 2016. IEEE.

[27] Song, G., H. Myeong, and K.M. Lee. Seednet: Automatic seed generation with deep reinforcement learning for robust interactive segmentation. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[28] Xu, N., et al. Deep interactive object selection. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[29] Li, Z., Q. Chen, and V. Koltun. Interactive image segmentation with latent diversity. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[30] Li, K. and W. Tao, Adaptive optimal shape prior for easy interactive object segmentation. IEEE Transactions on Multimedia, 2015. 17(7): p. 994-1005.

[31] Tang, M., et al. Grabcut in one cut. in Proceedings of the IEEE International Conference on Computer Vision. 2013.

[32] Yu, H., et al. Loosecut: Interactive image segmentation with loosely bounded boxes. in 2017 IEEE International Conference on Image Processing (ICIP). 2017. IEEE.

[33] Wu, S., M. Nakao, and T. Matsuda, SuperCut: Superpixel based foreground extraction with loose bounding boxes in one cutting. IEEE Signal Processing Letters, 2017. 24(12): p. 1803-1807.

[34] Cheng, M.-M., et al. Densecut: Densely connected crfs for realtime grabcut. in Computer Graphics Forum. 2015. Wiley Online Library.

[35] Andrade, F. and E.V. Carrera. Supervised evaluation of seed-based interactive image segmentation algorithms. in 2015 20th symposium on signal processing, images and computer vision (STSIVA). 2015. IEEE.

[36] Taha, A.A. and A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC medical imaging, 2015. 15(1): p. 1-28.

[37] Maninis, K.-K., et al. Deep extreme cut: From extreme points to object segmentation. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[38] Wang, T., et al., Diffusive likelihood for interactive image segmentation. Pattern Recognition, 2018. 79: p. 440-451.

[39] Chen, D.-J., H.-T. Chen, and L.-W. Chang, Toward a unified scheme for fast interactive segmentation. Journal of Visual Communication and Image Representation, 2018. 55: p. 393-403.

[40] Wang, D., et al., Region-based nonparametric model for interactive image segmentation. IEEE Access, 2019. 7: p. 111124-111134.

[41] Wang, T., et al., Probabilistic diffusion for interactive image segmentation. IEEE Transactions on Image Processing, 2018. 28(1): p. 330-342.

[42] Breve, F., Interactive image segmentation using label propagation through complex networks. Expert Systems With Applications, 2019. 123: p. 18-33.

[43] Dai, L., et al. Object extraction from bounding box prior with double sparse reconstruction. in Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015.

[44] Nguyen, T.N.A., et al., Robust interactive image segmentation using convex active contours. IEEE Transactions on Image Processing, 2012. 21(8): p. 3734-3743.

[45] Oh, C., B. Ham, and K. Sohn, Robust interactive image segmentation using structure-aware labeling. Expert Systems with Applications, 2017. 79: p. 90-100.

[46] Xian, M., et al., Neutro-connectedness cut. IEEE Transactions on Image Processing, 2016. 25(10): p. 4691-4703.

[47] Wang, T., et al., Global Manifold Learning for Interactive Image Segmentation. IEEE Transactions on Multimedia, 2021. 23: p. 3239-3249.

[48] Wang, T., et al., Interactive Image Segmentation Based on Label Pair Diffusion. IEEE Transactions on Industrial Informatics, 2020. 17(1): p. 135-146.

[49] Chai, S.S. and K.L. Goh, Evaluation of Different User Input Types in Interactive Segmentation. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 2017. 9(2-10): p. 91-99.

[50] Goh, K.L., et al., A Comparative Study of Interactive Segmentation with Different Number of Strokes on Complex Images. International Journal on Advanced Science, Engineering and Information Technology, 2020. 10: p. 178-184.

Design of Intelligent Fusion Terminal System with Fog Computing Capability in Distribution Area based on Large Capacity CPU

Ou Zhang[1]*, Songnan Liu[2], Hetian Ji[3], Xuefeng Wu[4], Xue Jiang[5]
State Grid Liaoning Electric Power Company Limited, Economic Research Institute, Shenyang, China[1, 4, 5]
Northeast Branch of State Grid Corporation of China, Shenyang, China[2]
State Grid Liaoning Marketing Service Center, Shenyang, China[3]

***Abstract*—The intelligent fusion terminal in the distribution area usually adopts the mode of cooperation between the cloud and the edge, and the workload of manual operation and maintenance is large. Therefore, an intelligent fusion terminal system in the distribution area with fog computing capability based on a high-capacity CPU is proposed. Follow the “cloud pipe edge end” construction framework of smart IOT system, and take this framework as the edge computing node of distribution station area and power consumption side. Mt7622b chip in Linux operating system with openwrt firmware is used as the main control chip of edge agent gateway equipment, and the recursive least square method is used to realize the data fusion of power acquisition service in distribution station area and power distribution demand terminal. The test results show that the designed system can realize real-time monitoring of power consumption and distribution data and power quality management in the distribution station area, and the data processing delay is less than 100ms, which provides a reference for the intelligent fusion terminal system in the distribution station area.**

Keywords—Large-capacity CPU; fog computing capability; power distribution station area; intelligent integration; terminal; system design

I. INTRODUCTION

The communication demands of power users and the intelligent terminal equipment of the power grid are growing linearly. The communication network of many terminal devices in the distribution network station area needs to have a high level of manual operation and maintenance, which puts forward higher requirements for delay, bandwidth and communication flexibility [1]. The intelligent terminal in the station area is the core unit of measurement and control management in the current intelligent power distribution system. The intelligent terminal in the station area is usually installed on the side of the distribution transformer. It uses the intelligent terminal in the station area to realize the communication interaction of the main station system and the access of the lower-level intelligent equipment [2-3]. The intelligent terminal equipment in the station area is of high importance and has a huge impact on the power distribution system. The distribution station area is the junction of production and marketing. Both the production and marketing parties install distribution transformer terminals and concentrators on the station side based on business needs. There are problems such as non-sharing of data, duplication of functions, and increased operation and maintenance workload [4-5]. The intelligent fusion terminal device can realize the functions of the concentrator and the distribution terminal, reducing the cost of equipment procurement, reducing the workload of manual operation and maintenance, and saving the cost of operation and maintenance.

At present, there is much researches on terminal data analysis of power distribution systems. Literature [6] designs a remote terminal data acquisition system for intelligent power distribution automation. The author in [7] proposes an automatically test system for intelligent substation relay protection equipment based on information fusion, which can separately collect data from remote terminals of intelligent power distribution and automatic test of intelligent substation relay protection equipment. Although the above research has achieved certain results and has certain application performance, it does not have the function of improving the data processing efficiency of the intelligent fusion terminal in the distribution station area. To this end, a large-capacity CPU-based intelligent fusion terminal system with fog computing capability is proposed in the distribution station area. Fog computing is based on the concept of cloud computing. In the fog computing mode, the data processing of many intelligent terminals is concentrated on the network edge devices, which effectively improves the data processing efficiency [8]. Applying fog computing to data management in distribution station area can effectively reduce data transmission delay and improve data communication capabilities. The innovation of the research is to use fog computing and large-capacity CPU to improve the data processing capability and communication performance of the system, as an intelligent terminal management device with many functions such as power quality monitoring, distribution transformer detection, humidity and temperature monitoring, etc. It has a high degree of integration, many functions, realizes remote control of intelligent terminals, and has high applicability.

*Corresponding Author.

II. Intelligent Fusion Terminal System in Distribution Station Area with Large-capacity CPU with Fog Computing Capability

A. System Overall Structure

The designed intelligent fusion terminal system in the distribution station area has the functions of a concentrator and TTU, and can be used as the edge computing node in the distribution station area and the power consumption side. The distribution station area takes the intelligent fusion terminal system of the distribution station area as the core, and establishes an intelligent low-voltage distribution Internet of Things with distribution automation. The overall structure of the intelligent fusion terminal system in the power distribution station area with a large-capacity CPU designed with fog computing capability is shown in Fig. 1.

As can be seen from Fig. 1, the designed system follows the "cloud-pipe-side-end" construction framework of the smart IoT system, which can realize the intelligent integration of the power distribution station area's use and procurement services and power distribution requirements. The system supports the communication protocols of the mining system, power distribution and unified IoT management center. Using container technology, multiple containers can be run at the same time, data communication between containers is supported, and data exchange and sharing functions between containers. The marketing application in the distribution station area and the distribution application are installed in their respective containers, and all the functions of the system are realized by using the application APP they have.

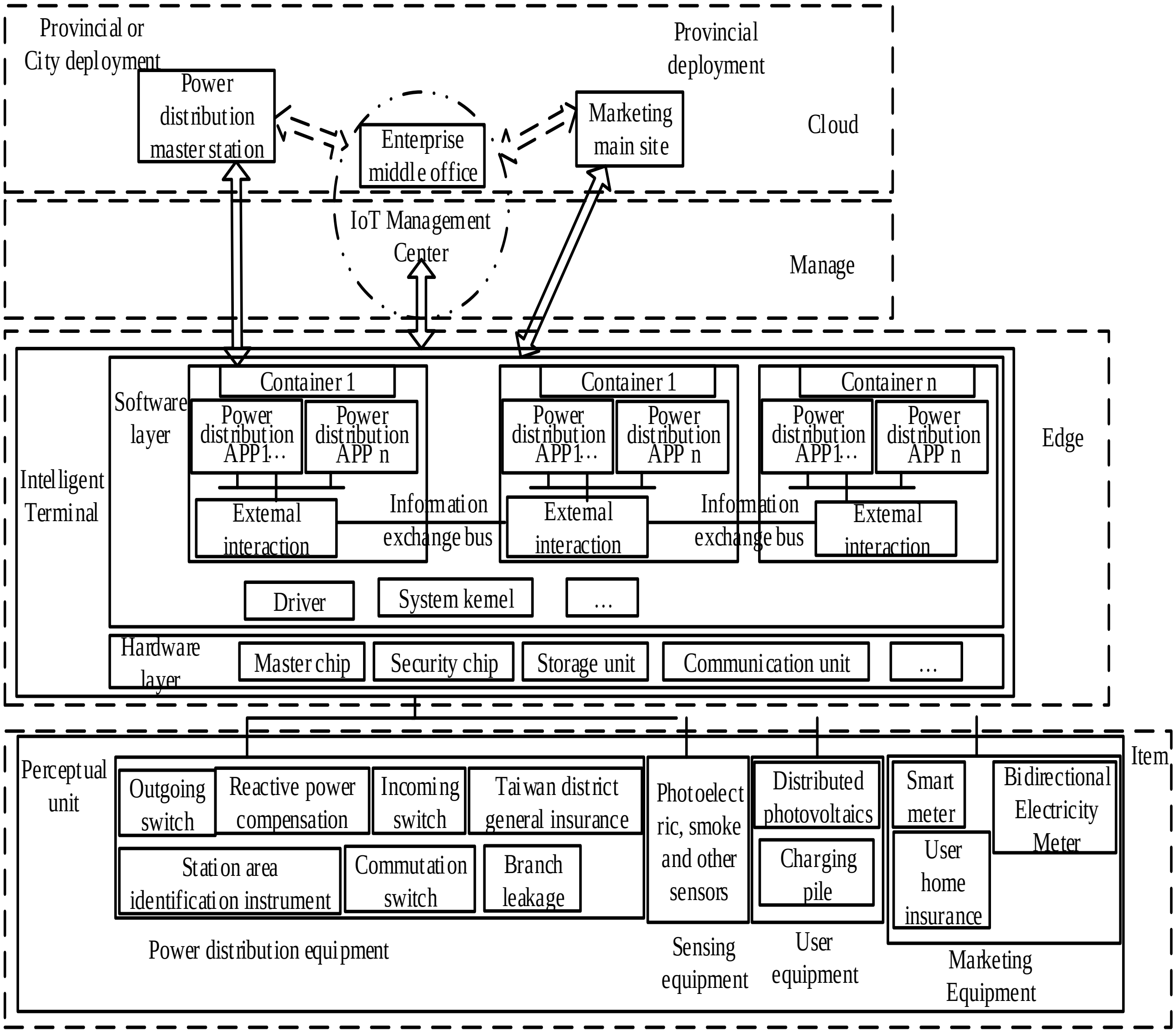


Fig. 1. System Overall Structure Diagram.

The system has external module communication data transmission application, container construction, data recording log and abnormal work log, running status monitoring, remote software and hardware information query, data recording and data statistics, event recording and reporting, parameter setting and query functions. The intelligent fusion terminal system has edge computing capabilities, and integrates the functions of power supply and consumption information collection in the distribution station area, energy meter or collection terminal data collection, equipment status monitoring and communication networking, local analysis and decision-making, and collaborative computing. The hardware adopts a platform-based design, supports the edge computing framework, and has fog computing capabilities [9], which can realize flexible expansion of functions through software-defined methods.

1) System hardware design: The hardware of the intelligent fusion terminal includes the basic core board, the main control board, the carrier module, the acquisition board, the power supply module, the single 4G module, and the dual 4G module. The hardware performance indicators of the fusion terminal are better than the reference standard requirements, and have good consistency. The hardware is designed based on the main control chip. The chip adopts the MT7622B architecture single-core 4-core processor, the main frequency is 1.35GHz, and the peripheral integrates 2GB DDR3 SDRAM and 8GM FLASH memory, two-way Gigabit Ethernet and other peripheral interfaces [10]. The software is based on an autonomous controllable operating system, Docker containers and autonomous orchestration technology.

The main control chip of the edge proxy gateway device adopts MT7622B, the peripheral is equipped with STM32 main control, and the MT7622B chip runs the OpenWRT operating system. It mainly realizes IO expansion and related acquisition functions. The software uses Ubuntu 16.04 operating system to compile openwrt firmware. Flash the compiled firmware into flash, and then run docker in the system to manage the entire application in a containerized manner. STM32 is developed with STM32 library to realize driver layer and application layer encapsulation.

The MT7622B chip adopts the Linux operating system of OpenWrt firmware. In order to ensure the normal and stable operation of the application, the firmware is cut and transplanted, and burned to the hardware platform. According to functional requirements, they increase or decrease related peripheral interface drivers, so that the overall system framework can meet the work requirements and be more streamlined. The main control chip has limited pins [11], and the terminal platform is equipped with an MCU chip as a data transfer hub. The MCU chip is used to connect the relevant peripheral modules, and communicate with the MT7622B main control chip through the communication interface, so that the main control chip can control the relevant peripheral modules.

MT7622B integrates dual ARM®Cortex-A53 cores, with a working frequency of up to 1.35GHz, with strong computing power and working stability. The high-speed DDR3 interface can be connected to high-capacity DRAM chips. It supports a variety of NAND Flash interfaces to realize mass storage control. The chip also includes various peripherals, including UART, SGMII, RGMII, PCIe2.0, USB2.0 (Host), USB3.0 (Host), and a 5-port 10/100 switch. It supports hardware router function, realizes 2.5Gbps HSGMII and 1Gbps RGMII Ethernet interface, embedded 5-port 10/100 switch and 802.11n 2.4GHz wireless. It can be supported by 802.11ac WLAN connection, and a variety of external network connection methods are available.

2) System software design: The system software platform provides a stable operating environment for the fusion terminal. The software platform is used to realize the decoupling of system software and hardware and software APP. The software platform includes the operating system and the edge computing framework.

The operating system completes the design of the root of trust and the chain of trust based on Linux. We have completed the driver design of all peripheral interfaces of the terminal, completed the kernel tailoring optimization and patch upgrade [12], and realized the kernel's support for the self-developed container management and terminal management platform.

The edge computing framework includes the container management engine Docker, system management, and various layers of components for edge-side application development, supporting various business applications. It has realized Docker's management of system-level single container and multiple APPs. It implements various functions such as container start, stop, uninstall, and deletion, APP start, stop, uninstall and deletion, and APP and container quota management.

B. Edge Proxy Gateway Device with Fog Computing Function

1) Edge proxy gateway device: Fog computing data processing and applications are concentrated in network edge devices rather than being stored in the cloud. The distribution station area is the intersection of production and marketing. Based on business needs, both production and marketing parties install distribution transformer terminals and concentrators on the side of the station area. There are problems such as non-sharing of data, duplication of functions, and increased operation and maintenance workload [13]. To solve such problems, the edge proxy gateway device with fog computing function is applied to the intelligent fusion terminal system in the distribution station area. The structure diagram of the edge proxy gateway device with fog computing function is shown in Fig. 2.

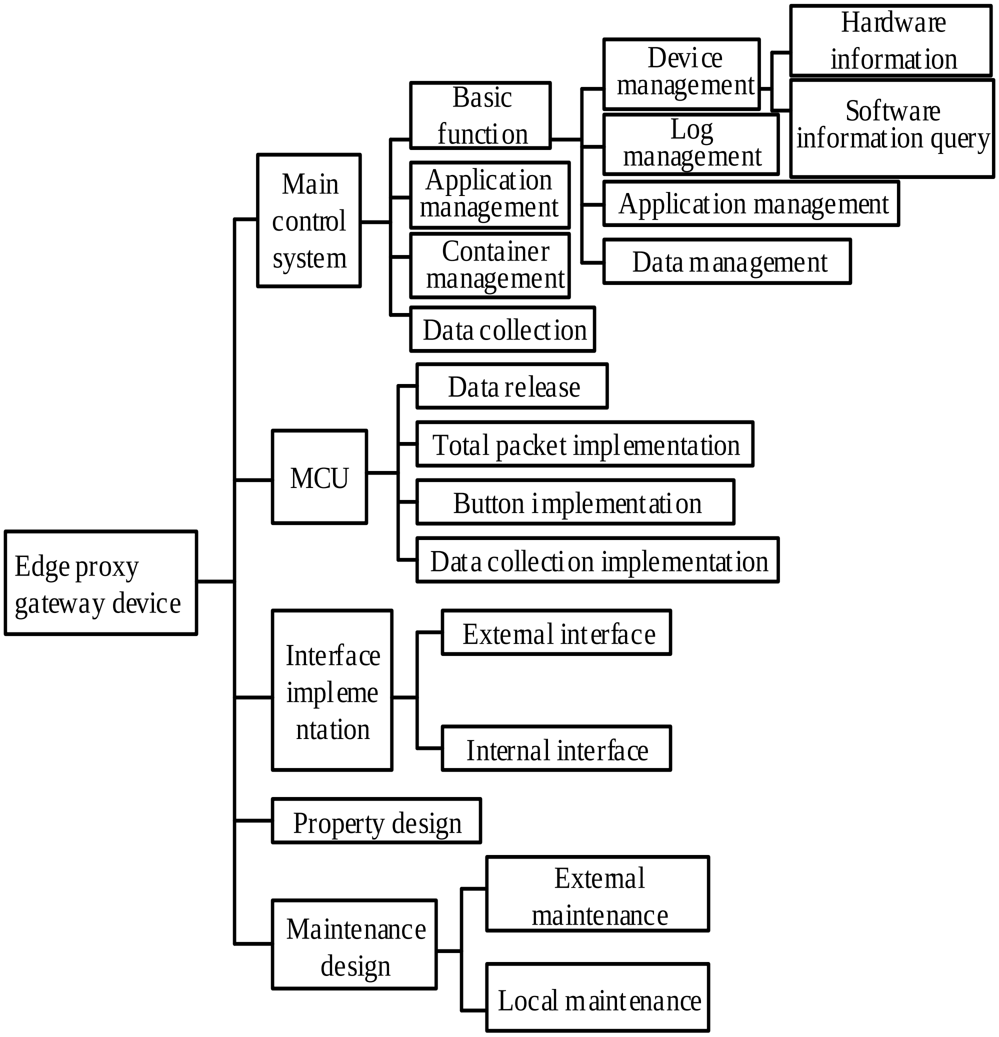


Fig. 2. Edge Proxy Gateway Device Structure Diagram.

The edge proxy gateway device includes a gateway system and a shell, and the gateway system includes a main control system, a single-chip microcomputer, an interface implementation, an attribute design and a system maintenance design. There is installation and positioning holes at the ends of the casing, and the installation and positioning holes are in a symmetrical state with respect to the central axis of the casing. A gas inlet window, a sequence area, a power button and a setting button are arranged on one side of the casing. The sequence area is located at the top of the gas inlet window, and the setting button is located between the sequence area and the power button. The edge proxy gateway device can monitor the devices in the distribution station area [14]. It can reduce the workload of equipment operation and maintenance, and reduce the input of manpower and material resources.

2) Network related skills reserve: The network-related skills reserves of the designed intelligent fusion terminal system in the distribution station area are as follows:

Layer 2 switch: Identify the MAC address information in the data packet. Forwarding is based on the MAC address, and the MAC address and the corresponding port are recorded in the internal address table.

Layer 3 switches: Layer 3 switches implement IP routing functions through a hardware switching mechanism. Its optimized routing software improves the efficiency of the routing process and solves the routing speed problem of traditional router software.

Link layer communication protocol: The data link layer provides services to the network layer based on the services provided by the physical layer. The most basic service is the reliable transmission of the source machine network layer data to the adjacent node target machine network layer.

Network layer communication protocol: The data link layer provides the data frame transmission function between two adjacent endpoints to manage network data communication. It transmits data from the source to the destination through several intermediate nodes, and provides end-to-end data transmission services to the transport layer [15].

DHCP: Dynamic Host Configuration Protocol (Dynamic Host Configuration Protocol), referred to as DHCP. It is a network protocol used in local area networks. This protocol allows servers to dynamically assign IP addresses and configuration information to clients.

DNS: Domain Name System (Domain Name System abbreviation DNS, Domain Name is translated as domain name) is the core service of the Internet. As a distributed database that can map domain names and IP addresses to each other, it makes Internet access more convenient without recording IP strings that can be read by machines.

LAN: Local Area Network (LAN) refers to a computer composed of multiple computers interconnected in a certain area.

WLAN: Wide Area Network (Wide Area Network), also known as a wide area network, external network or public network. The WAN is not the same as the Internet. It is a long-distance network that connects different local area networks or metropolitan area network computers to communicate, mainly using packet switching technology.

Docker container technology: A Linux container is a series of processes isolated from the rest of the system. All the files needed to run these processes are provided by another mirror. This means that Linux containers are portable and consistent from development to testing to production. Containers run faster than development pipelines that rely on repeating traditional test environments. Docker technology brings many new concepts and tools. It includes a simple command line interface to run and build new layered images, a server daemon, a library with prebuilt container images, and a registry server concept. Combining the above technologies, users can quickly build new layered containers and easily share containers with others.

3) Data fusion by recursive least squares: The edge computing layer of the system uses the recursive least squares method to realize the data fusion of the intelligent fusion terminal in the distribution station area. Assume that the number of smart terminals included in the distribution station area is n, and the measurement equation of smart terminal i is as follows:

$$Z_i = X + U_i \tag{1}$$

In equation (1), Z_i represents the i measurement value of the smart terminal. X represents the parameter value. U_i represents the noise value included in the data acquisition process.

The noise in the data collection process has no correlation, and the optimal weighting is based on a fixed criterion. According to the weight coefficient, the collected values of each intelligent terminal are multiplied to avoid the loss of information caused by removing the noise of data collection.

The weighting coefficient determination equation is as follows:

$$\hat{X} = \sum_{i=1}^{n} \alpha_i \times Z_i \tag{2}$$

In Equation (2), α_i represents a weighting coefficient.

The weighted estimated mean squared error equation is established as follows:

$$E = \sum_{i=1}^{n} \alpha_i^2 \delta_i^2 \times \hat{X} \tag{3}$$

In Equation (3), δ^2 represents the noise variance.

According to the conditional extreme value acquisition method of the above equation, the weighting coefficient equation that minimizes E is determined as follows:

$$f\left(\alpha_1, \alpha_2, \cdots, \alpha_n, \theta\right) = \sum_{i=1}^{n} \alpha_i^2 \delta_i^2 - \theta\left(\sum_{i=1}^{n} \alpha_i - 1\right) \tag{4}$$

In equation (4), θ represents the weight. The weighting coefficient equation is as follows:

$$\alpha_i = \frac{\delta_i^{-2}}{\sum_{i=1}^{n} \delta_i^{-2}} \tag{5}$$

The mean squared error equation to obtain the weighted estimate is as follows:

$$E\left[\left(X - \hat{X}\right)^2\right] = \frac{1}{\sum_{i=1}^{n} \delta_i^{-2}} \tag{6}$$

According to the determined optimal weights, the recursive relationship of the recursive least squares method is used to realize the data fusion of intelligent terminals in the distribution station area. The equation is as follows:

$$\begin{cases} P_{k+1} = P_k - P_k^2 / \left(1 + P_k\right) \\ \hat{Y}_{k+1} = \hat{Y}_k + P_{k+1}\left(\hat{X}_{k+1} - \hat{Y}_k\right) \end{cases} \tag{7}$$

In equation (7), $\hat{Y}_k$ and $\hat{Y}_{k+1}$ represent the data of the smart terminal at time k and $k+1$, respectively. Both P_k and P_{k+1} represent intermediate quantities. $\hat{X}_{k+1}$ represents the weighted estimated value of n intelligent terminals when time is $k+1$. Obtaining the fusion result by the recursive least squares method can ensure the minimum and unbiased estimated mean square error.

III. System Experimental Test

In order to verify the design of the large-capacity CPU with the ability of fog computing, the intelligent fusion terminal system management in the distribution station area and the effectiveness of monitoring the distribution station network. We use Matlab software to build the system in this paper, and select the distribution station area of a power company as the experimental object. The distribution station area includes distribution automation terminals, state monitoring intelligent terminals, power quality monitoring intelligent terminals, state monitoring sensors, secondary screen cabinets and many other intelligent terminals. In the distribution station area, the optical fiber network is used to realize communication.

After the system of this paper is adopted in the distribution station area, the emergency repair time when the fault occurs is calculated. The system in this paper is compared with the remote system [6] and the information fusion system [7]. The comparison results of the fault repair time are shown in Fig. 3.

It can be seen from the experimental results in Fig. 3 that the application of the system in this paper in the distribution station area can reduce the time of power failure and improve the reliability of power supply. The reliability of power supply in the distribution station area determines customer satisfaction and power supply revenue. After adopting the system in this paper, the dispatcher in the distribution station area can close the power outage switch at the first time in case of a fault, so as to reduce the outage time. It can notify the on-duty personnel to quickly repair, reduce the repair time, and improve the reliability of power supply.

At present, the residential voltage qualification rate cannot meet the residential electricity demand. The voltage qualification rate before the system in this distribution station area is not used is about 90%. After the system in this paper was applied to the distribution station area, the voltage quality of the distribution network from May to September 2019 was calculated. The method in this paper is compared with the remote system and the information fusion system. The comparison results are shown in Fig. 4.

It can be seen from the experimental results in Fig. 4 that after the system in this paper is adopted, the voltage qualification rate of the distribution station area is significantly improved. Within five months of operation, the system in this paper increased the voltage qualification rate of the distribution station area to more than 98.5%. The remote system and the information fusion system have a small improvement in the voltage qualification rate of the distribution station area, and the voltage qualification rate is lower than 95%. This system effectively monitors the voltage data of each integrated point. Through reasonable logical judgment, the comprehensive voltage regulation of the line voltage regulator is realized to ensure that the voltage of each node in the low-voltage line in the distribution network meets the voltage qualified value.

The real-time operation of the system is of great significance. The system in this paper adopts the fog computing architecture and has strong data processing and communication capabilities. Statistics The system in this paper is used to collect the refresh rate of the intelligent terminal equipment in the distribution station area. The statistical results are shown in Table I.

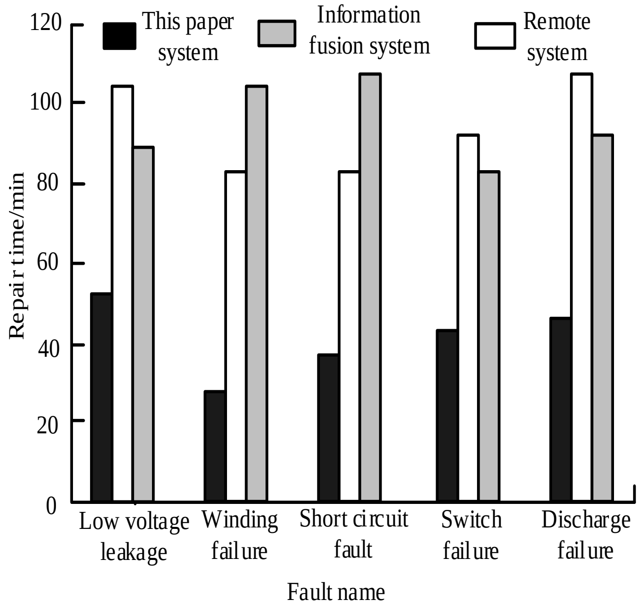


Fig. 3. Fault Repair Time Comparison.

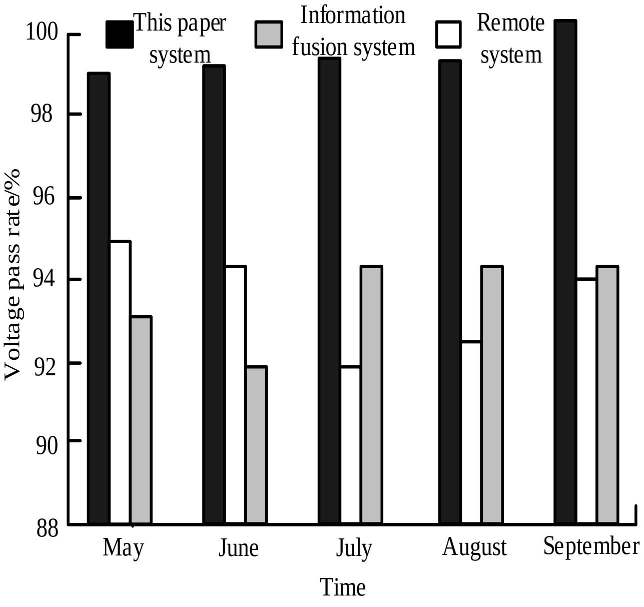


Fig. 4. Voltage Pass Rate Comparison.

TABLE I. Data Refresh Rate

System name	Fast data processing/(times/s)	Data collection/(times/s)
This paper system	120	30
Remote system	85	21
Information fusion system	94	19

It can be seen from the experimental results in Table I that the fast data refresh rate of the system in this paper is as high as 120 times/s. Intelligent terminal data collection refreshes data up to 30 times/s. The data refresh speed of the system in this paper is significantly higher than the other two systems, which verifies that the system in this paper has a higher data refresh speed. The system in this paper adopts fog computing architecture combined with large-capacity CPU, which has high communication performance and data transmission rate.

The monitoring time delay of the real-time monitoring of distribution data in the distribution station area is calculated using the system in this paper. The system in this paper is compared with the remote system and the information fusion system, and the comparison results are shown in Table II.

TABLE II. POWER DISTRIBUTION DATA MONITORING DELAY

Power distribution monitoring data name	This paper system/ms	Remote system/ms	Information fusion system/ms
Three-phase voltage	85	185	251
Electric current	91	165	234
Electrical energy	76	205	218
Tariff electricity	58	234	265
Total power	82	198	208
Voltage pass rate	94	152	234
Three-phase unbalance	84	184	218
Harmonic	76	235	269
Circuit breaker status	58	196	248
Capacitor switching state	64	184	235
Distribution terminal status	78	258	274

It can be seen from the experimental results in Table II that the monitoring delay of the system in this paper is less than 100ms. Using remote system and information fusion system to monitor the distribution of data monitoring delay in the distribution station area is higher than 150ms. The power distribution data monitoring delay of the system in this paper is significantly lower than that of the other two systems. The main reason is that the system in this paper adopts a fog computing architecture and a large-capacity CPU, which effectively improves the data processing capability of the system. The system meets the data acquisition and monitoring needs of the distribution station area.

Statistical data processing efficiency under different data volumes: The system in this paper is compared with the other two systems, and the comparison results are shown in Fig. 5.

As can be seen from the experimental results in Fig. 5, the data processing efficiency of the system in this paper is higher than 97% under different data volumes. The data processing efficiency of the other two systems is lower than 96% in the case of different data volumes. The fog computing architecture adopted in this system has high data processing efficiency. The previous cloud computing platform needs to transmit all the collected data to the computing center, and use the computing center to process the data, which cannot meet the delay requirement. The fog computing architecture can realize the analysis of data collected by the edge computing unit and improve the efficiency of data processing. The fog computing architecture can meet the low latency requirements of power communication networks. In the face of massive power data, it has high processing efficiency and can be used as a good solution in the distribution network.

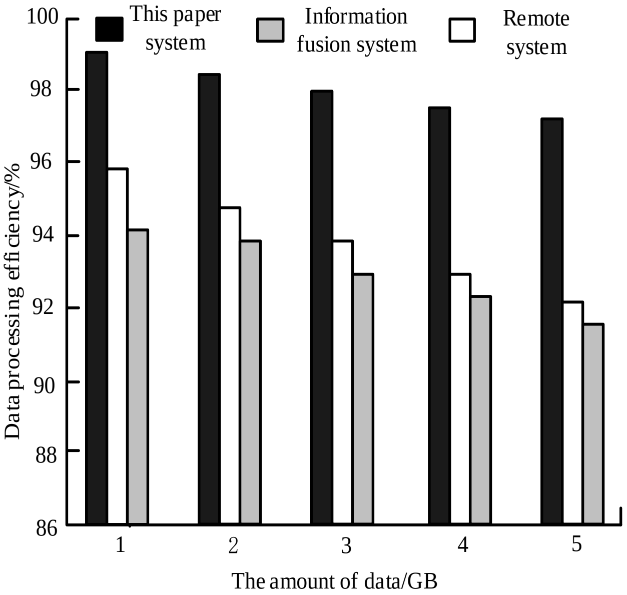


Fig. 5. Data Processing Efficiency Comparison Results.

After adopting the system in this paper, the cost of investing in secondary equipment in the distribution station area is calculated. The system in this paper is compared with the other two systems, and the comparison results are shown in Table III.

TABLE III. SECONDARY EQUIPMENT INVESTMENT ANALYSIS

Project	This paper system/million	Remote system/million	Information fusion system/million
Distribution automation terminal	3.5	4.2	4.3
Power quality monitoring terminal	2.4	2.6	2.5
Condition monitoring sensors	1.9	2.2	2.3
Condition monitoring unit	1.8	1.9	2.1
Fiber optic communication unit	0.9	1.1	1.1
Secondary screen cabinet	0.7	0.9	0.8
Total	11.2	12.9	13.1

From the experimental results in Table III, it can be seen that after using the system in this paper, the cost of investing in secondary equipment in the distribution station area is lower than the 17,000 yuan of the remote system and the 19,000 yuan of the information fusion system. The experimental results show that the system in this paper has a higher management effect of power distribution equipment. Applying the system in this paper to the distribution station area can effectively save the investment cost. It has high performance of intelligent terminal management in power distribution station area, which is helpful for the long-term development of electric power enterprises.

IV. Conclusion

Power grid technology has developed rapidly, and power infrastructure has been gradually improved. The primary and secondary equipment of the distribution network has gradually matured and stabilized. There are many intelligent terminals installed in the distribution network station area. Through the intelligent terminals, the power distribution data collection in the distribution station area and the online monitoring of the power consumption data can be realized. However, the functions of comprehensive management, comprehensive analysis and comprehensive control in the distribution station area are not perfect. We combined the fog computing mechanism with the large-capacity CPU to design an intelligent fusion terminal system in the distribution station area with fog computing capability based on the large-capacity CPU. And it is verified by experiments that the system has high practicability in the distribution station area. The researched system has a high level of automation and low input cost, which can help power companies to complete the goal of smart grid construction and realize smart grid functions. In the event of a power outage in the distribution station area, the researched system can transmit the terminal fault information data back to the system. The system sends information to equipment maintenance personnel. After receiving the information, the maintenance personnel can quickly reach the fault location, and the repair speed is fast, which improves customer satisfaction with electricity consumption.

V. Declaration

The authors declare that there are no conflicts of interest regarding the publication of this paper.

VI. Funding

The study was supported by “2018 quota station technology project of State Grid Corporation <Research on billing basis of UAV aerial survey technology application> (Grant NO. SGDE-LN-2018-02-003)” and “Science and technology project of State Grid Corporation <Research on monitoring and control optimization of power grid life cycle investment> (Grant NO. 2015YF-64)”.

References

[1] Liu D Q, Zeng X J, Wang Y N. Security Situation Assessment of Intelligent Distribution Transformer Terminal Unit Based on Information Entropy[J]. Southern Power System Technology, 2020, 121(01):24-29.

[2] Ji Q H, Wei Z, Wang T, et al. Research on the Control Strategy for Three-phase Unbalanced Load in Distribution Area[J]. Power Electronics, 2020, 327(02):38-40+52.

[3] Li Y X, Lu J, Xu Z Q, et al. Design of Terminal Communication Access Architecture for Smart Power Distribution and Utilization Based on Integration of Multiple Technologies[J]. Automation of Electric Power Systems, 2018, 632(10):169-175.

[4] Liu T, Tang L, He X Q, et al. Optimal Task Offloading Scheme Based on Network Delay and Resource Management in Joint Blockchain and Fog Computing System[J]. Journal of Electronics & Information Technology, 2020, 42(09):2180-2185.

[5] Wang C L, Huang X T. Study on Optimal Allocation of Inference Nodes for Fog Computing in Smart Environment[J]. Acta Electronica Sinica, 2020, 048(001):35-43.

[6] Qi E J, Peng D G, Guan X L, et al. Design of the Remote Terminal Data Acquisition System for the Intelligent Distribution Automation[J]. Electric Drive, 2018, 345(07):62-67.

[7] Chen F J, Lv Y S, Fan G S, et al. Automatic test system of relay protection device for smart substation based on information fusion technology[J]. Power System Protection and Control, 2020, 551(05):164-169.

[8] Cong W, Sheng Y R, Xian G F, et al. Distributed Power Service Restoration Method Based on Smart Terminal Unit[J]. Automation of Electric Power Systems, 2018, 637(15):83-91.

[9] Fu J, Ji R Y, Wang D, et al. Research and Application of Distribution Network Courts User' 's Behavior Analysis Model Based on Parallel K-Means Clustering[J]. Power System and Clean Energy 2018, 034(011):71-76.

[10] Ma W, Shao C X, Liu M Z. Study on content access of fog computing in highway service area[J]. Application of Electronic Technique, 2018, 44(012):101-105, 110.

[11] Liu L, Zhao G Q. Intelligent Incentive Mechanism for Fog Computing-based Multimedia Systems with Swarming Behavior[J]. Computer Science, 2019, 046(011):94-99.

[12] Ou Z Z, Zhao H J, Zhu L J, et al. Research on the Method for Three-phase Load Balancing and Reactive Power Compensation in Distribution Areas[J]. Power Electronics, 2019, 317(04):101-104.

[13] Xiong J Y, Yang G P, Guo Y, et al. Compatibility Analysis of ECT and EVT in Primary & Secondary Fusion of Intelligent Power Distribution Equipment[J]. Power Capacitor & Reactive Power Compensation, 2019, 40(01):114-120.

[14] Liu R L, Liu H T, Xia S F, et al. Internet of Things Technology Application and Prospects in Distribution Transformer Service Area Management[J]. High Voltage Engineering, 2019, 319(06):33-40.

[15] Wu S, Hao S P, Yang L X, et al. Information model and application of intelligent distribution area based on CIM[J]. Electrical Measurement & Instrumentation, 2018, 55(10):52-57+95.

A Blind Robust Image Watermarking on Selected DCT Coefficients for Copyright Protection

Majid Rahardi[1], Ferian Fauzi Abdulloh[2], Wahyu Sukestyastama Putra[3]
Faculty of Computer Science, Universitas Amikom Yogyakarta
Sleman, Indonesia[1, 2, 3]

***Abstract*—This paper proposes a blind and robust image watermarking technique using Discrete Cosine Transform (DCT) for copyright protection on color images called BRIW-DCT. Each channel of the host image is divided into non-overlapping image blocks with the size of 8×8 pixels. Each image block is transformed into a frequency domain using the DCT transformation. The watermark image is embedded into the host image by modifying the 11th to the 15th DCT coefficient. The experimental result shows that the watermarked image achieved a high PSNR value of 50.4489 dB and a high SSIM value of 0.9991. Furthermore, various attacks are performed on the watermarked image. BRIW-DCT can successfully recover the watermark image from the tampered image, which produces a high NC value of 0.7805 and a low BER value of 0.1126.**

Keywords*—Robust watermarking; copyright protection; discrete cosine transform; frequency domain; color image watermarking

I. Introduction

The development of internet technology in recent years has contributed to the birth of social media platforms. The advancement of the mobile operating system such as Android and iOS also contributes to the growth of social media users [1]–[4]. Everyone can share their data seamlessly across many social media platforms. One type of data is a multimedia image. The multimedia image is created using a camera and various available editing software [5]–[11]. The image itself is then uploaded to the social media platform. Everyone can download and re-upload the image on another platform. This action may violate the copyright and ownership of the image. In order to protect the ownership of the image, some artist commonly adds a visible watermark to the image. However, someone who has skill in image editing software may remove or replace the watermark on the image. As a result, the owner of the image has lost the intellectual property of that image. To solve this problem, researchers have developed an invisible watermark to protect the ownership of the image [12]–[18].

There are three categories of invisible image watermarking: robust, semi-fragile, and fragile image watermarking. A fragile image watermarking scheme embeds the watermark image into the host image in the original domain of the image, which is the spatial domain. The watermark embedding is performed on the Least Significant Bit (LSB) of the image. While semi-fragile and robust watermarking embed the watermark data in the transform domain, such as Discrete Cosine Transform (DCT) [19]–[28], Discrete Wavelet Transform (DWT) [29]–[37], and Singular Value Decomposition (SVD) [38]–[46]. Based on the objective, fragile and semi-fragile watermarking is commonly utilized for image authentication. In contrast, robust watermarking is widely used for copyright protection. In image authentication, the embedded watermark data must be sensitive to any modification to the image. Thus, if an image area is modified, the technique can localize the tampered area [47]–[50]. In contrast, in copyright protection, the watermark data must be preserved for any modification to the image, such as image compression [51]–[60]. Thus, it is called robust watermarking.

The watermarking process consists of two steps: embedding and extracting the watermark data [61]. The owner of the digital image does watermark embedding the first time the image is created. The image itself is then uploaded to the internet. If someone steals and modifies that image, then the owner can prove his ownership through watermark extraction from the modified image [62]. The extracted watermark then can be used as evidence in the court for justice. There are three techniques in the watermark extraction process: semi-blind, blind, and non-blind watermarking [63]. The non-blind technique requires the information from the host image and the watermark logo in the extraction process. The semi-blind technique requires additional information, such as the embedding region coordinates. In contrast, the blind technique does not require any information from the host image. Thus, the blind technique is the most efficient method in the robust image watermarking.

This paper proposes a blind and robust image watermarking scheme primarily used for copyright protection, namely BRIW-DCT. At first, the scheme divides the host image into three RGB channels. Each channel is divided into non-overlapping image blocks with the size of 8×8 pixels. Each block is then transformed into the frequency domain using the DCT. Each pixel of the watermark data is embedded into each block by modifying the 11th to the 15th DCT coefficient. The selected embedding location is considered the optimum for embedding without corrupting the host image. Once the watermark data is embedded, the inverse DCT is performed to reveal the watermarked image. The watermarked image can then be distributed safely through the internet. The evaluation of the watermarked image is computed using Structural Similarity Index Measure (SSIM) and Peak Signal to Noise Ratio (PSNR). While the extraction of the watermark data is measured using Normalized Cross-Correlation (NC) and Bit Error Rate (BER).

The rest of this article is organized as follows: Section II presents the related works of the existing robust watermarking techniques. The proposed method is explained in Section III. The experimental result and analysis are shown in Section IV. Finally, Section V concludes this research.

II. Related Work

Yousevi et al. [12] presented a blind robust image watermarking scheme on color images using the Integer Wavelet Transform (IWT). The scheme divided the color image into non-overlapping image blocks with the size of 4×4 pixels. Next, each block was transformed into a frequency domain using IWT. The low sub-band is selected as the embedding location of the watermark data to improve the watermarked image quality. The watermark data is embedded in a chaotic manner using the Lyapunov exponent. This process prevents the illegal extraction of the watermark data. Furthermore, the chaotic map is randomized using the Pseudo-Random Number Generator (PRNG). The experiments apply various attacks to the watermarked image, such as salt and pepper, low pass filtering, cropping, blurring, etc. The experimental result shows that the scheme successfully embeds the watermark data into the host image. However, the extracted watermark doesn't reach a satisfactory level of imperceptibility. Thus, the technique can be improved further.

Zermi et al. [13] presented a robust digital watermarking scheme using DWT and SVD for medical images. The host image is transformed into a frequency domain using DWT. Furthermore, the LL sub-band is transformed using SVD. The watermark data is then embedded into the SVD coefficient matrix. The watermark data itself is generated from the electronic patient record. The watermark data is then formatted into a binary sequence of data and hashed using the MD5 function. The experiment was performed using the Ocular Disease Intelligent Recognition (ODIR) database. The images are tampered with using various tampering methods such as JPEG compression, average filtering, gamma correction, sharpening, and scaling. The experimental results showed that the scheme could maintain the imperceptibility of the watermarked image. The scheme was also robust against several conventional attacks. However, the scheme has the limitation of usage on the medical images. The scheme can be further improved to support various types of multimedia images.

Begum et al. [14] presented a hybrid and robust watermarking scheme using DCT, DWT, and SVD. The Arnold map was used to encrypt the watermark image. The host image was transformed in the frequency domain using DCT followed by DWT and finalized using SVD. The experiment was conducted using various tampering attacks such as median filter and rotation attacks. The experimental result shows that the scheme achieved high robustness against multiple attacks. However, the utilization of two transform domains led to high computational costs. Thus, the scheme can further be improved to reduce the computational cost while maintaining robustness.

Fares et al. [15] presented a blind robust image watermarking based on the Fourier transform. Fourier transform is the first introduced frequency domain transformation in signal processing research. The scheme separated the color images into each RGB component. The Fourier transform is applied individually on each channel. Furthermore, multiple variants of the Fourier transform were utilized. Those variants are Fractional Fourier Transform (FFT), Quaternion Discrete Fourier Transform (QDFT), and Discrete Fourier Transform (DFT). The watermark image was inserted into the selected coefficient of the Fourier Transform. Once the watermark data was embedded, the inverse transformation was performed to produce the watermarked image. The experiment was done using multiple attack scenarios such as histogram equalization, blurring, rescaling, Gaussian noise, rotation, and JPEG compression. The experimental results showed that the scheme successfully embedded the watermark data into the host image. In addition, the scheme could also extract the watermark data under various attack scenarios. However, the extracted watermark quality can still be improved further.

Laxmanika and Singh [16] presented a robust image watermarking scheme using DWT, SVD, DCT, Particle Swarm Optimization (PSO), and Bi-dimensional Empirical Mode Decomposition (BEMD). The host image is decomposed using 2nd level DWT into sub-bands. The selected bands were then decomposed further using the BEMD. To optimize the searching of complex multidimensional data, the PSO was implemented. Furthermore, the DCT followed by SVD is applied to the selected band. In his research, the security key was utilized in the embedding process. The extraction process extracted the watermark data in reverse. The experimental result showed that the scheme was robust in restoring the watermark data after various attacks were applied to the watermarked image. However, excessive use of multiple transform domains led to a high computational cost. Therefore, the scheme can be further improved to reduce the computational time.

Thanki et al. [17] presented a blind watermarking scheme using Discrete Curvelet Transform (DCuT) and Redundant Discrete Wavelet Transform (RDWT). It combined two transformation domains to improve the imperceptibility of the watermarked image. A hybrid coefficient is selected from a single-level RDWT and the high-frequency DCuT. At first, the scheme implemented DCuT. The scheme then took the high-frequency coefficient and transformed it into RDWT. The watermark data was embedded into the LH sub-band of RDWT. The scheme also implemented Arnold Transform and Pseudo-random Noise (PN) sequences to scramble the watermark data. The scheme implemented multiple scaling factors between 5 and 40. In a lower scaling factor value, the scheme produced a high imperceptibility. However, the robustness was sacrificed. In contrast, the scheme achieved high robustness on a high scaling factor value while sacrificing imperceptibility. In addition, utilizing multiple transform domains has contributed to high computational complexity, reducing the watermark embedding speed. Thus, the scheme can be improved further.

Abdulrahman and Ozturk [18] presented a robust color image watermarking using DCT and DWT transformation. The DCT and DWT were applied to each of the RGB components. The scheme also used Arnold Transform to scramble the watermark data from a grayscale watermark image. Various image processing attacks are applied to the image, such as filtering, JPEG compression, resizing, and rotating. The experimental result has shown that the scheme can produce a high imperceptibility on a low scaling factor and high robustness on high scaling factors. However, the dual-domain approach has contributed to high computational costs. Hence, improvements are required.

III. Proposed Method

Eight color images from the SIPI-USC image database are utilized as the dataset for this research. University of Southern California (USC) provided this image for image processing research. Each image has a size of 512×512 pixels. Furthermore, many researchers used these images to experiment in the image watermarking field. The images are shown in Fig. 1.

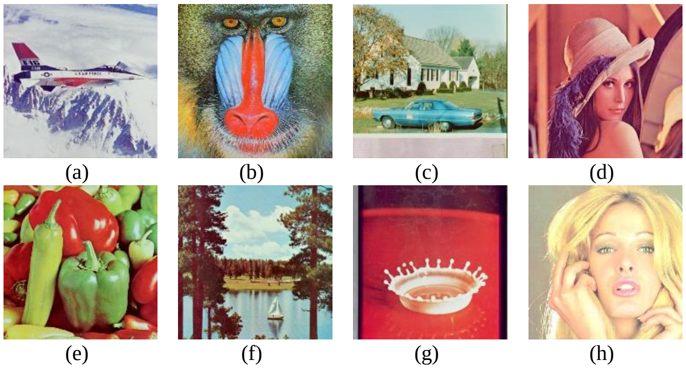


Fig. 1. The Host Images (a) Airplane (b) Baboon (c) House (d) Lena (e) Peppers (f) Sailboat (g) Splash (h) Tiffany.

A. Watermark Embedding

The scheme embeds the watermark data into the host images. Each of the host images in Fig. 1 will undergo the watermark embedding process, as visualized in Fig. 2.

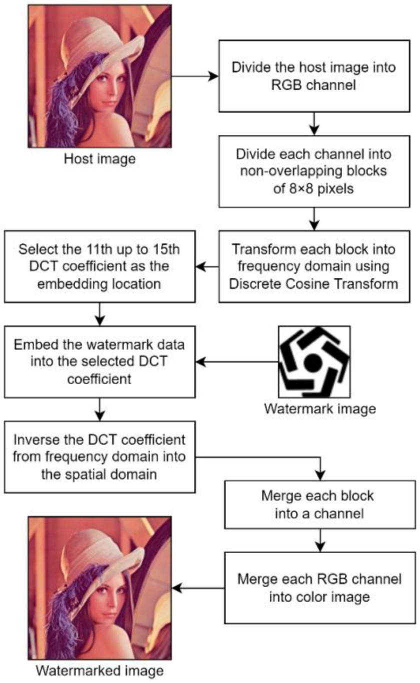

Host image
Divide the host image into RGB channel
Divide each channel into non-overlapping blocks of 8×8 pixels
Transform each block into frequency domain using Discrete Cosine Transform
Select the 11th up to 15th DCT coefficient as the embedding location
Embed the watermark data into the selected DCT coefficient
Watermark image
Inverse the DCT coefficient from frequency domain into the spatial domain
Merge each block into a channel
Merge each RGB channel into color image
Watermarked image

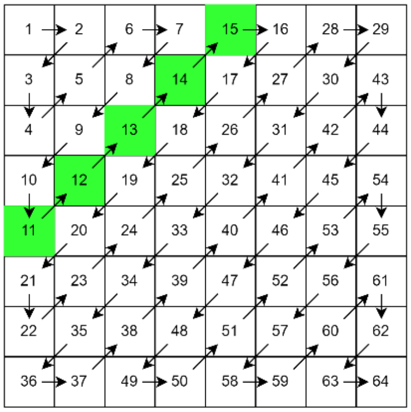
Fig. 2. BRIW-DCT Watermark Embedding Process.

According to Fig. 2, the embedding watermark process starts from the host image divided into RGB channels. Each channel is divided into non-overlapping image blocks with the size of 8×8 pixels. Next, each block is transformed using DCT into a frequency domain. The watermark data is then embedded into the selected DCT coefficient. The watermark itself is taken from the logo of Universitas Amikom Yogyakarta. The watermark image is stored in the binary black-and-white image with the size of 64×64 pixels. There are 64 coefficients for each block, as visualized in Fig. 3.

BRIW-DCT embeds the watermark data into the DCT coefficient, which has a low frequency between the 11th to the 15th. The purpose of the low-frequency selected DCT coefficient is to ensure the robustness of BRIW-DCT. Once the watermark data is embedded, the DCT coefficient is inverted into the spatial domain. Each block is then merged into a channel. And each channel is merged into the watermarked image. The watermark embedding process is also explained in Algorithm 1.

Algorithm 1. BRIW-DCT watermark embedding algorithm

```
Input: host, watermark
    [height, width, channel] = size(host);
    blockSize = 8;
    blockHeight = ceil(height / blockSize);
    blockWidth = ceil(width / blockSize);
    watermarked = zeros(height, width, channel, 'uint8');
    for y = 1:blockHeight
        yMax = y * blockSize;
        yMin = yMax - blockSize + 1;
        for x = 1:blockWidth
            xMax = x * blockSize;
            xMin = xMax - blockSize + 1;
            block = host(yMin:yMax, xMin:xMax, :);
            watermarked(yMin:yMax, xMin:xMax, :) =
            embedBlock(block, watermark(y, x));
        end
    end
    function output = embedBlock(input, wm)
        [~, ~, channel] = size(input);
        scale = 4;
        output = input;
        for c = 1:channel
            block = input(:, :, c);
            dct = dct2(block);
            dct(1, 5) = writeWm(dct(1, 5), scale, wm);
            dct(2, 4) = writeWm(dct(2, 4), scale, wm);
            dct(3, 3) = writeWm(dct(3, 3), scale, wm);
            dct(4, 2) = writeWm(dct(4, 2), scale, wm);
            dct(5, 1) = writeWm(dct(5, 1), scale, wm);
            idct = uint8(idct2(dct));
            output(:, :, c) = idct;
        end
    end
    function output = writeWm(input, scale, wm)
        base = (fix(input / scale) * scale);
        offset = (wm * scale / 2);
        output = base + offset;
    end
Output: watermarked
```



1	2	6	7	15	16	28	29
3	5	8	14	17	27	30	43
4	9	13	18	26	31	42	44
10	12	19	25	32	41	45	54
11	20	24	33	40	46	53	55
21	23	34	39	47	52	56	61
22	35	38	48	51	57	60	62
36	37	49	50	58	59	63	64

Fig. 3. Selected DCT Coefficients for Embedding.

According to Algorithm 1, the input of the algorithm is the host image and the watermark image. The output of the algorithm is the watermarked image. The watermark embedding is defined in the embedBlock() function. The DCT coefficient modification is defined in writeWm() function. The DCT transformation and inversion process are shown in Line 22 and 28, respectively. The watermark data is embedded in the five selected DCT coefficients on each RGB component for redundancy. Thus, if one watermark is broken, another watermark data can be extracted.

B. *Watermark Extraction*

Once the watermark data is successfully embedded into the host image, the watermarked image is ready to be distributed on the internet safely. If the image is misused and modified by an unauthorized user, the actual owner of the image can perform the extraction process to reveal the watermark data. The extraction process is explained in Fig. 4.

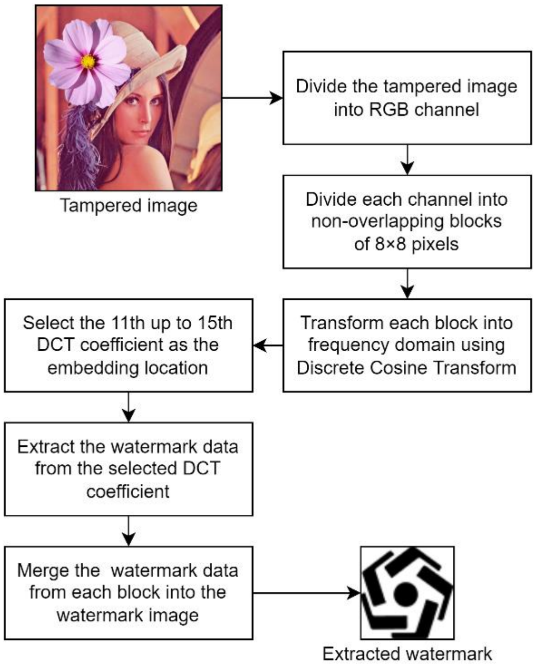

Tampered image
Divide the tampered image into RGB channel
Divide each channel into non-overlapping blocks of 8×8 pixels
Transform each block into frequency domain using Discrete Cosine Transform
Select the 11th up to 15th DCT coefficient as the embedding location
Extract the watermark data from the selected DCT coefficient
Merge the watermark data from each block into the watermark image
Extracted watermark

Fig. 4. BRIW-DCT Watermark Extraction Process.

Based on Fig. 4, the tampered image is divided into RGB channels. Each channel is then divided into non-overlapping blocks of 8×8 pixels. The tampered image is then transformed using DCT into the frequency domain. The scheme then selects the 11th up to 15th DCT coefficient of each block to extract the watermark bit. The watermark bit of each block is merged into 64×64 pixels of watermark data, producing a binary watermark image. The extraction process of the watermark image is also explained in Algorithm 2.

Algorithm 2: The watermark extraction algorithm

```
Input: tampered
1     [height, width, channel] = size(tampered);
2     blockSize = 8;
3     blockHeight = ceil(height / blockSize);
4     blockWidth = ceil(width / blockSize);
5     watermark = zeros(blockHeight, blockWidth, 'logical');
6     for y = 1:blockHeight
7      yMax = y * blockSize;
8      yMin = yMax - blockSize + 1;
9      for x = 1:blockWidth
10     xMax = x * blockSize;
11     xMin = xMax - blockSize + 1;
12     block = tampered(yMin:yMax, xMin:xMax, :);
13     watermark(yMin:yMax, xMin:xMax, :) = extractBlock(block);
14     end
15    end
16    function output = extractBlock (input)
17     [~, ~, channel] = size(input);
18     scale = 4;
19     out = zeros(channel, 1, 'logical');
20     for c = 1:channel
21     wm = zeros(1, 3, 'logical');
22     block = input(:, :, c);
23     dct = dct2(block);
24     wm(1) = readWm(dct(1, 5), scale);
25     wm(2) = readWm(dct(2, 4), scale);
26     wm(3) = readWm(dct(3, 3), scale);
27     wm(4) = readWm(dct(4, 2), scale);
28     wm(5) = readWm(dct(5, 1), scale);
29     out(c) = nnz(wm == 1) > 2;
30     end
31     output = nnz(out == 1) > 1;
32    end
33    function output = readWm (input, scale)
34     coef = mod(input, scale);
35     limit = scale / 4;
36     output = limit < coef && coef < limit * 3;
37    end
Output: watermark
```

Based on Algorithm 2, the input of the algorithm is the tampered image. The output of the algorithm is the watermark image. The watermark extraction is defined in the extractBlock() function. The selected DCT coefficient extraction process is defined in readWm() function. The DCT transformation process is shown in Line 23.

C. *Performance Evaluation*

In order to measure the imperceptibility of the watermarked image, the scheme computes the PSNR and SSIM of the watermarked image. A high PSNR and SSIM value represents an insignificant difference between the host and watermarked images. Both measurements are commonly used in the field of image watermarking. The PSNR is defined by:

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(p(i,j) - q(i,j)\right)^2 \quad (1)$$

$$PSNR = 10\, log_{10}\left(\frac{255^2}{MSE}\right) \quad (2)$$

where p represents the host image, q represents the watermarked image, i and j represent the pixel coordinates. The PSNR values are represented in decibel (dB). Typically, the human visual system cannot distinguish two images with a PSNR value above 40 dB. The SSIM is defined by:

$$SSIM(i,j) = [l(i,j)]^{\alpha} \cdot [c(i,j)]^{\beta} \cdot [s(i,j)]^{\gamma} \quad (3)$$

$$l(p,q) = \frac{2\mu_p\mu_q + D_1}{\mu_p^2 + \mu_q^2 + D_1} \quad (4)$$

$$c(p,q) = \frac{2\sigma_p\sigma_q + D_2}{\sigma_p^2 + \sigma_q^2 + D_2} \quad (5)$$

$$s(p,q) = \frac{\sigma_{pq} + D_3}{\sigma_p\sigma_q + D_3} \quad (6)$$

where *l* is the luminance function to measure the closeness of the luminance of two images, *c* is the function of contrast to compute the contrast similarity of two images, *s* is the function of the structure to calculate the correlation coefficient between two images, D_1, D_2, and D_3 are constants with positive values. The SSIM utilizes the human visual system to measure the similarity between two images. Thus, it is considered more accurate compared to PSNR measurement. The robustness is measured using Bit Error Rate (BER) and Normalized Cross-Correlation (NC). The BER and NC are defined by:

$$BER = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} w(i,j) \oplus e(i,j)}{M \times N} \quad (7)$$

$$NC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} w(i,j).e(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} w(i,j)^2 \; \sum_{i=1}^{M}\sum_{j=1}^{N} e(i,j)^2}} \quad (8)$$

where *w* denotes the actual watermark image, *e* represents the watermark image that has been extracted from the tampered image. *M* and *N* denote the height and the width of the watermark image. A high NC value means the extracted watermark image is highly correlated to the actual watermark. While a high BER value denotes that the extracted watermark image has a higher error value than the actual watermark image, which frequently occurs when the image is under attack.

IV. Results and Analysis

The experiment in this research is evaluated using a computer with a 1.8 GHz octa-core AMD Ryzen 7 5700U processor, 32 GB memory, and a Windows 10 Home operating system. This experiment uses MATLAB 2021a as the programming language.

A. *The Performance of Imperceptibility*

In the process of watermark embedding, the watermark image is embedded in the DCT transformation domain. Thus, the watermarked image has invisible distortion when compared to the host image. The watermarked image and the host image of Lena are shown in Fig. 5.

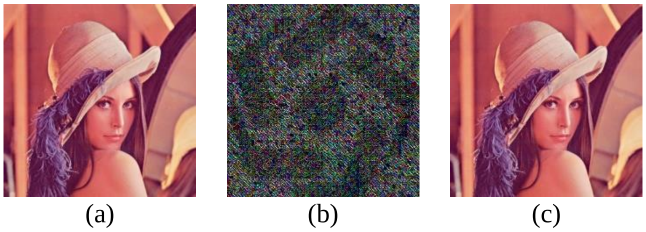


Fig. 5. The Lena Image (a) The Host Image (b) The Delta Image (c) The Watermarked Image.

According to Fig. 5, the scheme successfully embeds the watermark into the host image, as shown in Fig. 5(c). In addition, the watermark image is visually imperceptible from the watermarked image. The difference will only be visible if the delta image is brightened and sharpened into multiple levels, as shown in Fig. 4(b). Otherwise, the difference between the host and watermarked images is invisible to the human visual system. The watermarked image is evaluated using SSIM and PSNR. A higher SSIM and PSNR value means the watermarked image has less distortion compared to the actual image. On the other hand, a lower PSNR and SSIM value means the watermarked image suffers a significant error distortion compared to the host image. In addition, the computational time is also presented in this paper. It is expected to provide a complete picture of the performance of BRIW-DCT. The comparison of the imperceptibility is shown in Table I.

According to Table I, BRIW-DCT can maintain the watermarked image quality. The average watermarked image PSNR value is 50.4489 dB, while the average SSIM value is 0.9991. It proves that BRIW-DCT produces high imperceptibility in the watermark embedding process. In addition, BRIW-DCT requires less than one second to embed the watermark data. In order to show its full potential performance, this paper also compares the imperceptibility between BRIW-DCT and the existing scheme with similar watermarking techniques. Previously, Yousefi et al. [12] performed various experiments to protect the copyright of images. The scheme implemented various transform domains such as IWT, DWT, and CT. The result showed that the DWT method performed better in terms of PSNR and execution time. In terms of SSIM, the CT method has the highest imperceptibility. Another scheme by Thanki et al. [17] achieved the lowest imperceptibility in terms of PSNR value. The scheme also has the highest computational time to embed the watermark data due to implementing the dual-domain approach. However, the scheme by Thanki et al. [17] has a slightly better SSIM value than Yousefi et al. scheme [12]. The imperceptibility comparison with related work is presented in Table II.

Based on Table II, BRIW-DCT outperforms the previous method in terms of imperceptibility under PSNR and SSIM metrics. BRIW-DCT takes slightly more time to embed the watermark data. However, the execution time is highly dependent on the computer specification, which has a possibility of slight variation in the embedding speed. A computer with a higher clock rate and memory size can execute faster than the lower one. Thus, the execution time cannot be utilized as the main comparison between schemes.

TABLE I. THE IMPERCEPTIBILITY COMPARISON OF BRIW-DCT BETWEEN IMAGES

Image	PSNR (dB)	SSIM	Time (s)
Airplane	50.5074	0.9963	0.6719
Baboon	49.9688	0.9996	0.6094
House	50.5124	0.9989	0.7500
Lena	50.3421	0.9997	0.9219
Pepper	50.3113	0.9997	0.6719
Sailboat	50.1928	0.9993	0.6094
Splash	50.8296	0.9995	0.7188
Tiffany	50.9268	0.9996	0.6719
Average	50.4489	0.9991	0.7032

TABLE II. THE IMPERCEPTIBILITY COMPARISON WITH RELATED WORK

Method	PSNR (dB)	SSIM	Time (s)
Yousefi et al. (IWT) [12]	48.8236	0.9880	0.6643
Yousefi et al. (DWT) [12]	49.8228	0.9890	**0.6196**
Yousefi et al. (CT) [12]	48.8221	0.9925	0.6877
Thanki et al. (DCuT & RDWT) [17]	47.6514	0.9953	1.5734
Proposed BRIW-DCT	**50.4489**	**0.9991**	0.7032

B. The Performance of Robustness

Various attack scenarios are implemented into the watermarked image to compute the robustness of BRIW-DCT. The watermark data is then extracted from the tampered image. At first, the watermarked image is tampered with using various tampering attacks. The attack scenarios are presented in Table III.

Table III shows multiple attack scenarios on the subject images to show the robustness of BRIW-DCT. The Airplane image is not modified with any tampering attack as the control. The Baboon image is modified using the Gaussian filter. The Gaussian noise is the most common attack applied to the image. The House image is modified using the salt & pepper noise. The Lena image is sharpened using the sharpen filter. The Pepper image is modified using median filtering. The sailboat image is attacked using the ripple mask. The splash image is modified using the mosaic filter. Finally, the tiffany image is modified using the unsharp filter. The tampered image quality is calculated using the PSNR and SSIM measurement against the host image. The tampered image has an average PSNR value of 36.3787 dB and an average SSIM value of 0.9771. The extracted watermark image is then compared to the actual watermark image. The extracted watermark image is shown in Fig. 6.

Based on Fig. 6, the mosaic filter in 6g dramatically affects the quality of the extracted watermark image. In contrast, the unsharp filter in 6h produces a less significant effect on the image. The Airplane image in 6a, which was used as the control image, can completely recover the watermark data, proven by the NC value of 1 and BER value of 0. Overall, the embedded watermark logo can be preserved under various tampering attacks. It proves that BRIW-DCT is robust in maintaining the watermark logo for copyright protection. The robustness comparison between images is presented in Table IV.

Table IV shows that the robustness varies between the images. It is highly affected by the type and the severity of the tampering attack. The average BER and NC values are 0.1226 and 0.7805, respectively. It proves the robustness of BRIW-DCT in the watermark extraction process. Furthermore, BRIW-DCT can extract the watermark image in under half a second, enabling it to implement in mobile devices with low computational power.

TABLE III. THE ATTACK SCENARIOS

Image	Attack	PSNR (dB)	SSIM
Airplane	No Attack	50.5074	0.9963
Baboon	Gaussian Filter	26.7028	0.9256
House	Salt & Pepper	32.5146	0.9382
Lena	Sharpen	41.1576	0.9980
Pepper	Median Filter	35.7176	0.9950
Sailboat	Ripple Mask	33.6158	0.9887
Splash	Mosaic Filter	30.7983	0.9777
Tiffany	Unsharp Filter	40.0151	0.9973
Average		36.3787	0.9771

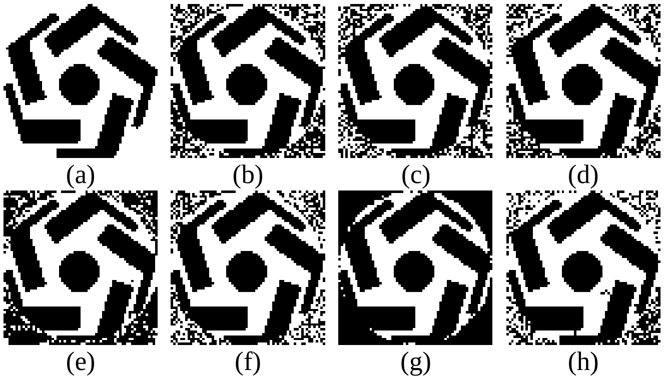


Fig. 6. The Extracted Watermark Image (a) Airplane (b) Baboon (c) House (d) Lena (e) Peppers (f) Sailboat (g) Splash (h) Tiffany.

TABLE IV. THE ROBUSTNESS COMPARISON BETWEEN IMAGES

Image	Attack	NC	BER	Time (s)
Airplane	No Attack	1.0000	0.0000	0.3594
Baboon	Gaussian Filter	0.7454	0.1396	0.3125
House	Salt & Pepper	0.7667	0.1265	0.3281
Lena	Sharpen	0.7939	0.1121	0.2969
Pepper	Median Filter	0.6879	0.1772	0.2969
Sailboat	Ripple Mask	0.7913	0.1123	0.3281
Splash	Mosaic Filter	0.6386	0.2158	0.3125
Tiffany	Unsharp Filter	0.8200	0.0972	0.3125
Average		0.7805	0.1226	0.3184

V. Conclusion

This paper presented a blind and robust technique for color image watermarking based on DCT for copyright protection. Each image’s block has been transformed into the transform domain using the DCT. The watermark data has been embedded into the host image by modifying the 11th up to the 15th DCT coefficients. The experimental results conducted in this research have shown that the watermarked image achieved a high PSNR value of 50.4489 dB and a high SSIM value of 0.9991. Various attacks have been applied to the watermarked image to show the performance of BRIW-DCT. It shows that BRIW-DCT can achieve a high NC value of 0.7805 and a low BER value of 0.1126. In the future, BRIW-DCT can be improved by implementing the Arnold Transform to enhance the robustness against image tampering attacks.

Acknowledgment

This research was supported by the Research Department of Universitas Amikom Yogyakarta

References

[1] A. Aminuddin, “Android Assets Protection Using RSA and AES Cryptography to Prevent App Piracy,” 2020 3rd Int. Conf. Inf. Commun. Technol. ICOIACT 2020, pp. 461–465, Nov. 2020, doi: 10.1109/ICOIACT50329.2020.9331988.

[2] F. Ernawan, N. A. Abu, and H. Rahmalan, “Tchebichef moment transform on image dithering for mobile applications,” Fourth Int. Conf. Digit. Image Process. (ICDIP 2012), vol. 8334, pp. 83340D-83340D–5, May 2012, doi: 10.1117/12.946023.

[3] A. Sukma Darmawan et al., “Tree-based Ensemble Learning for Stress Detection by Typing Behavior on Smartphones,” Proc. - 2021 Int. Conf. Softw. Eng. Comput. Syst. 4th Int. Conf. Comput. Sci. Inf. Manag. ICSECS-ICOCSIM 2021, pp. 394–398, Aug. 2021, doi: 10.1109/ICSECS52883.2021.00078.

[4] M. S. Bin Othman Mustafa, M. Nomani Kabir, F. Ernawan, and W. Jing, “An Enhanced Model for Increasing Awareness of Vocational Students Against Phishing Attacks,” 2019 IEEE Int. Conf. Autom. Control Intell. Syst. I2CACIS 2019 - Proc., pp. 10–14, Jun. 2019, doi: 10.1109/I2CACIS.2019.8825070.

[5] Z. Mustaffa, M. H. Sulaiman, B. Yusob, and F. Ernawan, “Integration of GWO-LSSVM for time series predictive analysis,” IET Conf. Publ., vol. 2016, no. CP688, 2016, doi: 10.1049/CP.2016.1360.

[6] Z. Mustaffa, M. H. Sulaiman, D. Rohidin, F. Ernawan, and S. Kasim, “Time series predictive analysis based on hybridization of meta-heuristic algorithms,” Int. J. Adv. Sci. Eng. Inf. Technol., vol. 8, no. 5, pp. 1919–1925, 2018, doi: 10.18517/IJASEIT.8.5.4968.

[7] I. Khandokar, M. Hasan, F. Ernawan, S. Islam, and M. N. Kabir, “Handwritten character recognition using convolutional neural network,” J. Phys. Conf. Ser., vol. 1918, no. 4, Jun. 2021, doi: 10.1088/1742-6596/1918/4/042152.

[8] L. J. Halawa, A. Wibowo, and F. Ernawan, “Face Recognition Using Faster R-CNN with Inception-V2 Architecture for CCTV Camera,” ICICOS 2019 - 3rd Int. Conf. Informatics Comput. Sci. Accel. Informatics Comput. Res. Smarter Soc. Era Ind. 4.0, Proc., Oct. 2019, doi: 10.1109/ICICOS48119.2019.8982383.

[9] L. Kartikawati, I. Nabawi, V. Rahayu, Kusrini, D. Ariatmanto, and F. Ernawan, “Physical Distancing System Using Computer Vision,” 3rd Int. Conf. Cybern. Intell. Syst. ICORIS 2021, 2021, doi: 10.1109/ICORIS52787.2021.9649548.

[10] M. L. Prasetyo et al., “Face Recognition Using the Convolutional Neural Network for Barrier Gate System,” Int. J. Interact. Mob. Technol., vol. 15, no. 10, pp. 138–153, 2021, doi: 10.3991/IJIM.V15I10.20175.

[11] A. Mujaddidurrahman, F. Ernawan, A. Wibowo, E. A. Sarwoko, A. Sugiharto, and M. D. R. Wahyudi, “Speech Emotion Recognition Using 2D-CNN with Data Augmentation,” Proc. - 2021 Int. Conf. Softw. Eng. Comput. Syst. 4th Int. Conf. Comput. Sci. Inf. Manag. ICSECS-ICOCSIM 2021, pp. 685–689, Aug. 2021, doi: 10.1109/ICSECS52883.2021.00130.

[12] M. Yousefi Valandar, M. Jafari Barani, and P. Ayubi, “A blind and robust color images watermarking method based on block transform and secured by modified 3-dimensional Hénon map,” Soft Comput., vol. 24, no. 2, pp. 771–794, Nov. 2019, doi: 10.1007/S00500-019-04524-Z.

[13] N. Zermi, A. Khaldi, R. Kafi, F. Kahlessenane, and S. Euschi, “A DWT-SVD based robust digital watermarking for medical image security,” Forensic Sci. Int., vol. 320, p. 110691, Mar. 2021, doi: 10.1016/J.FORSCIINT.2021.110691.

[14] M. Begum, J. Ferdush, and M. Shorif Uddin, “A Hybrid robust watermarking system based on discrete cosine transform, discrete wavelet transform, and singular value decomposition,” J. King Saud Univ. - Comput. Inf. Sci., Jul. 2021, doi: 10.1016/J.JKSUCI.2021.07.012.

[15] K. Fares, K. Amine, and E. Salah, “A robust blind color image watermarking based on Fourier transform domain,” Optik (Stuttg)., vol. 208, p. 164562, Apr. 2020, doi: 10.1016/J.IJLEO.2020.164562.

[16] Laxmanika and P. K. Singh, “Robust and imperceptible image watermarking technique based on SVD, DCT, BEMD and PSO in wavelet domain,” Multimed. Tools Appl., pp. 1–26, Aug. 2021, doi: 10.1007/S11042-021-11246-8.

[17] R. Thanki, A. Kothari, and D. Trivedi, “Hybrid and blind watermarking scheme in DCuT – RDWT domain,” J. Inf. Secur. Appl., vol. 46, pp. 231–249, Jun. 2019, doi: 10.1016/J.JISA.2019.03.017.

[18] A. K. Abdulrahman and S. Ozturk, “A novel hybrid DCT and DWT based robust watermarking algorithm for color images,” Multimed. Tools Appl., vol. 78, no. 12, pp. 17027–17049, Jan. 2019, doi: 10.1007/S11042-018-7085-Z.

[19] F. Ernawan and M. N. Kabir, “A Robust Image Watermarking Technique With an Optimal DCT-Psychovisual Threshold,” IEEE Access, vol. 6, pp. 20464–20480, Mar. 2018, doi: 10.1109/ACCESS.2018.2819424.

[20] F. Ernawan, M. Ramalingam, A. S. Sadiq, and Z. Mustaffa, “An improved imperceptibility and robustness of 4×4 DCT-SVD image watermarking using modified entropy,” J. Telecommun. Electron. Comput. Eng., vol. 9, no. 2–7, pp. 111–116, 2017.

[21] M. Fuad and F. Ernawan, “Video steganography based on DCT psychovisual and object motion,” Bull. Electr. Eng. Informatics, vol. 9, no. 3, pp. 1015–1023, Jun. 2020, doi: 10.11591/eei.v9i3.1859.

[22] F. Ernawan, M. N. Kabir, and Z. Mustaffa, “A blind watermarking technique based on DCT psychovisual threshold for a robust copyright protection,” 2017 12th Int. Conf. Internet Technol. Secur. Trans. ICITST 2017, pp. 92–97, May 2018, doi: 10.23919/ICITST.2017.8356354.

[23] D. Ariatmanto and F. Ernawan, “Adaptive scaling factors based on the impact of selected DCT coefficients for image watermarking,” J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 3, pp. 605–614, Mar. 2022, doi: 10.1016/J.JKSUCI.2020.02.005.

[24] D. Ariatmanto and F. Ernawan, “An adaptive scaling factor for multiple watermarking scheme,” 2019 4th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2019, pp. 175–178, Nov. 2019, doi: 10.1109/ICITISEE48480.2019.9003948.

[25] F. Ernawan, D. Ariatmanto, Z. Musa, Z. Mustaffa, and J. M. Zain, “An Improved Robust Watermarking Scheme using Flexible Scaling Factor,” 2020 Int. Conf. Comput. Intell. ICCI 2020, pp. 266–269, Oct. 2020, doi: 10.1109/ICCI51257.2020.9247798.

[26] F. Ernawan and M. N. Kabir, “An Improved Watermarking Technique for Copyright Protection Based on Tchebichef Moments,” IEEE Access, vol. 7, pp. 151985–152003, 2019, doi: 10.1109/ACCESS.2019.2948086.

[27] F. Ernawan, “Robust image watermarking based on psychovisual threshold,” J. ICT Res. Appl., vol. 10, no. 3, pp. 228–242, 2016, doi: 10.5614/ITBJ.ICT.RES.APPL.2016.10.3.3.

[28] N. A. Abu, F. Ernawan, N. Suryana, and S. Sahib, “Image watermarking using psychovisual threshold over the edge,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 7804 LNCS, pp. 519–527, 2013, doi: 10.1007/978-3-642-36818-9_60.

[29] F. Ernawan, D. Ariatmanto, and A. Firdaus, “An Improved Image Watermarking by Modifying Selected DWT-DCT Coefficients,” IEEE Access, vol. 9, pp. 45474–45485, 2021, doi:

10.1109/ACCESS.2021.3067245.

[30] F. Ernawan, S. C. Liew, Z. Mustaffa, and K. Moorthy, "A blind multiple watermarks based on human visual characteristics," Int. J. Electr. Comput. Eng., vol. 8, no. 4, pp. 2578–2587, Aug. 2018, doi: 10.11591/IJECE.V8I4.PP2578-2587.

[31] N. A. Abu, F. Ernawan, and N. Suryana, "An image dithering via Tchebichef moment transform," J. Comput. Sci., vol. 9, no. 7, pp. 811–820, 2013, doi: 10.3844/JCSSP.2013.811.820.

[32] H. Rahmalan, F. Ernawan, and N. A. Abu, "Tchebichef Moment Transform for colour image dithering," ICIAS 2012 - 2012 4th Int. Conf. Intell. Adv. Syst. A Conf. World Eng. Sci. Technol. Congr. - Conf. Proc., vol. 2, pp. 866–871, 2012, doi: 10.1109/ICIAS.2012.6306136.

[33] P. W. Adi, F. Ernawan, A. Wibowo, E. A. Sarwoko, and F. Agung Nugroho, "Watermarking Scheme based on Chinese Remainder Theorem and Integer Wavelet Filters for Copyright Protection," Proc. - 2021 Int. Conf. Softw. Eng. Comput. Syst. 4th Int. Conf. Comput. Sci. Inf. Manag. ICSECS-ICOCSIM 2021, pp. 70–74, Aug. 2021, doi: 10.1109/ICSECS52883.2021.00020.

[34] F. Ernawan, M. N. Kabir, M. Fadli, and Z. Mustaffa, "Block-based Tchebichef image watermarking scheme using psychovisual threshold," Proc. - 2016 2nd Int. Conf. Sci. Technol. ICST 2016, pp. 6–10, Mar. 2017, doi: 10.1109/ICSTC.2016.7877339.

[35] N. A. Abu, F. Ernawan, and F. Salim, "Smooth Formant Peak Via Discrete Tchebichef Transform," J. Comput. Sci., vol. 11, no. 2, pp. 351–360, 2015, doi: 10.3844/JCSSP.2015.351.360.

[36] M. Fuad, F. Ernawan, and L. J. Hui, "Video scene change detection based on histogram analysis for hiding message," J. Phys. Conf. Ser., vol. 1918, no. 4, Jun. 2021, doi: 10.1088/1742-6596/1918/4/042141.

[37] M. Fuad and F. Ernawan, "Frames selection based on modified entropy and object motion in video steganography," Int. J. Sci. Technol. Res., vol. 8, no. 10, pp. 761–766, Oct. 2019.

[38] F. Ernawan and M. N. Kabir, "A block-based RDWT-SVD image watermarking method using human visual system characteristics," Vis. Comput., vol. 36, no. 1, pp. 19–37, Jan. 2020, doi: 10.1007/S00371-018-1567-X.

[39] N. Alias and F. Ernawan, "Multiple watermarking technique based on rdwt-svd and human visual characteristics," J. Theor. Appl. Inf. Technol., vol. 97, no. 14, pp. 3980–3989, 2019.

[40] F. Ernawan and D. Ariatmanto, "Image watermarking based on integer wavelet transform-singular value decomposition with variance pixels," Int. J. Electr. Comput. Eng., vol. 9, no. 3, pp. 2185–2195, Jun. 2019, doi: 10.11591/IJECE.V9I3.PP2185-2195.

[41] F. Ernawan and M. N. Kabir, "A blind watermarking technique using redundant wavelet transform for copyright protection," Proc. - 2018 IEEE 14th Int. Colloq. Signal Process. its Appl. CSPA 2018, pp. 221–226, May 2018, doi: 10.1109/CSPA.2018.8368716.

[42] N. Alias and F. Ernawan, "Multiple watermarking technique using optimal threshold," Indones. J. Electr. Eng. Comput. Sci., vol. 18, no. 1, pp. 368–376, 2019, doi: 10.11591/IJEECS.V18.I1.PP368-376.

[43] D. Ariatmanto and F. Ernawan, "An improved robust image watermarking by using different embedding strengths," Multimed. Tools Appl., vol. 79, no. 17, pp. 12041–12067, Jan. 2020, doi: 10.1007/S11042-019-08338-X.

[44] F. Ernawan, "Tchebichef image watermarking along the edge using YCoCg-R color space for copyright protection," Int. J. Electr. Comput. Eng., vol. 9, no. 3, pp. 1850–1860, Jun. 2019, doi: 10.11591/IJECE.V9I3.PP1850-1860.

[45] F. Ernawan, P. W. Adi, S. C. Liew, E. A. Sarwoko, and E. Winarno, "Fast image watermarking based on signum of cosine matrix," Indones. J. Electr. Eng. Comput. Sci., vol. 25, no. 3, pp. 1383–1391, Mar. 2022, doi: 10.11591/IJEECS.V25.I3.PP1383-1391.

[46] F. Ernawan and M. F. Abdullah, "A New Embedding Technique Based on Psychovisual Threshold for Robust and Secure Compressed Video Steganography," 2020 33rd Gen. Assem. Sci. Symp. Int. Union Radio Sci. URSI GASS 2020, Aug. 2020, doi: 10.23919/URSIGASS49373.2020.9231989.

[47] A. Aminuddin and F. Ernawan, "AuSR1: Authentication and self-recovery using a new image inpainting technique with LSB shifting in fragile image watermarking," J. King Saud Univ. - Comput. Inf. Sci., Feb. 2022, doi: 10.1016/J.JKSUCI.2022.02.009.

[48] K. S. Lian, L. S. Chuin, and F. Ernawan, "Reversible Face Watermarking Scheme using Hash Function for Tamper Localization and Recovery," Proc. - 2021 Int. Conf. Softw. Eng. Comput. Syst. 4th Int. Conf. Comput. Sci. Inf. Manag. ICSECS-ICOCSIM 2021, pp. 58–63, Aug. 2021, doi: 10.1109/ICSECS52883.2021.00018.

[49] J. T. Lei Lei, L. S. Chuin, and F. Ernawan, "An Image Watermarking based on Multi-level Authentication for Quick Response Code," Proc. - 2021 Int. Conf. Softw. Eng. Comput. Syst. 4th Int. Conf. Comput. Sci. Inf. Manag. ICSECS-ICOCSIM 2021, pp. 417–422, Aug. 2021, doi: 10.1109/ICSECS52883.2021.00082.

[50] F. Ernawan, A. Aminuddin, D. Nincarean, M. F. A. Razak, and A. Firdaus, "Three Layer Authentications with a Spiral Block Mapping to Prove Authenticity in Medical Images," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 4, 2022, doi: 10.14569/IJACSA.2022.0130425.

[51] F. Ernawan, M. N. Kabir, Z. Mustaffa, K. Moorthy, and M. Ramalingam, "An Improved Image Compression Technique using Large Adaptive DCT Psychovisual Thresholds," Proc. 2nd IEEE Int. Conf. Knowl. Innov. Invent. 2019, ICKII 2019, pp. 561–564, Jul. 2019, doi: 10.1109/ICKII46306.2019.9042705.

[52] N. A. Abu and F. Ernawan, "A novel psychovisual threshold on large DCT for image compression," Sci. World J., vol. 2015, 2015, doi: 10.1155/2015/821497.

[53] F. Ernawan, N. A. Abu, and N. Suryana, "Adaptive tchebichef moment transform image compression using psychovisual model," J. Comput. Sci., vol. 9, no. 6, pp. 716–725, 2013, doi: 10.3844/JCSSP.2013.716.725.

[54] F. Ernawan, E. Noersasongko, and N. A. Abu, "An efficient 2×2 Tchebichef moments for mobile image compression," 2011 Int. Symp. Intell. Signal Process. Commun. Syst. "The Decad. Intell. Green Signal Process. Commun. ISPACS 2011, 2011, doi: 10.1109/ISPACS.2011.6146066.

[55] F. Ernawan, N. A. Abu, and N. Suryana, "An adaptive JPEG image compression using psychovisual model," Adv. Sci. Lett., vol. 20, no. 1, pp. 26–31, Jan. 2014, doi: 10.1166/ASL.2014.5255.

[56] F. Ernawan, N. A. Abu, and N. Suryana, "TMT quantization table generation based on psychovisual threshold for image compression," 2013 Int. Conf. Inf. Commun. Technol. ICoICT 2013, pp. 202–207, 2013, doi: 10.1109/ICOICT.2013.6574574.

[57] F. Ernawan, N. Kabir, and K. Z. Zamli, "An efficient image compression technique using Tchebichef bit allocation," Optik (Stuttg)., vol. 148, pp. 106–119, Nov. 2017, doi: 10.1016/J.IJLEO.2017.08.007.

[58] N. A. Abu, F. Ernawan, and N. Suryana, "A generic psychovisual error threshold for the quantization table generation on JPEG image compression," Proc. - 2013 IEEE 9th Int. Colloq. Signal Process. its Appl. CSPA 2013, pp. 39–43, 2013, doi: 10.1109/CSPA.2013.6530010.

[59] F. Ernawan, N. A. Abu, and N. Suryana, "An optimal tchebichef moment quantization using psychovisual threshold for image compression," Adv. Sci. Lett., vol. 20, no. 1, pp. 70–74, Jan. 2014, doi: 10.1166/ASL.2014.5316.

[60] F. Ernawan and S. H. Nugraini, "The optimal quantization matrices for jpeg image compression from psychovisual threshold," J. Theor. Appl. Inf. Technol., vol. 70, no. 3, pp. 566–572, 2014.

[61] R. Rajkumar and A. Vasuki, "Reversible and robust image watermarking based on histogram shifting," Cluster Comput., vol. 22, no. 5, pp. 12313–12323, Sep. 2019, Accessed: May 09, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s10586-017-1614-9.

[62] J. Abraham and V. Paul, "An imperceptible spatial domain color image watermarking scheme," J. King Saud Univ. - Comput. Inf. Sci., vol. 31, no. 1, pp. 125–133, Jan. 2019, doi: 10.1016/J.JKSUCI.2016.12.004.

[63] R. Thanki and S. Borra, "Fragile watermarking for copyright authentication and tamper detection of medical images using compressive sensing (CS) based encryption and contourlet domain processing," Multimed. Tools Appl., vol. 78, no. 10, pp. 13905–13924, May 2019, doi: 10.1007/s11042-018-6746-2.

Feedback Model when Applying the Evaluation by Indicators in the Development of Competences through Problem based Learning in a Systems Engineering Course

César Baluarte-Araya, Oscar Ramirez-Valdez
Departamento Académico de Ingeniería de Sistemas e Informática
Universidad Nacional de San Agustín de Arequipa
Arequipa, Perú

***Abstract*—Feedback can be very influential in students' learning, therefore, the university must be very clear about its procedures and rules on the time lapse of the response of the work done by them, and the comments made to positively influence in a sustained manner in the evaluation of learning. The present work shows the experience of applying the Problem Based Learning (PBL) methodology and also developing research competencies through Formative Research, and as a result of the evaluation of the learning of the Criteria and its Performance Indicators corresponding to the course Business Electronic which is taught by two teachers in theory and laboratory practices. The objective is to design a Feedback Model for the problems solved by the students in order to support the improvement of their learning. The methodology used is Problem Based Learning together with the Feedback Model, of real problems posed contemplating different contexts of the organisations; we have that from the Deliverable Report of each problem at the same time the incidences and observations are registered in the corresponding register and in this way the Feedback Report is elaborated. The results obtained reveal that the objectives of producing the Feedback Report are achieved, which should be sent as soon as possible to the students for analysis, to propose their own strategies for improving the shortcomings or errors, as well as having the motivation to continue progressing by accepting the suggestions or contributions of the teacher; as well as seeing an increase in knowledge, development of their competences, skills, attitudes, making their own judgements, and achieving the Student Results. In conclusion, the application of a well-planned active didactic strategy, the adequate evaluation of learning through the qualification of the indicators of each criterion, and the elaboration of a timely feedback report on the problems, will achieve the expected results for both the course and the student.**

Keywords*—*Problem based learning; competencies; evaluation; criteria; performance indicator; deliverable report; feedback report; skills; formative inquiry

I. Introduction

The Universidad Nacional de San Agustín de Arequipa [1], Arequipa - Peru, its Educational Model is based on professional training based on student competences, eliminating deficiencies in knowledge, soft skills and attitudes in students, and being able to use didactic strategies [2],[3],[4], adapting them to the nature of the same to achieve the training objectives and competences.

In 2021, the Professional School of Systems Engineering (EPIS) [5] underwent the evaluation process for accreditation by the Accreditation Board for Engineering and Technology (ABET) [6], one of its objectives being that the educational institution demonstrates that its graduates and those in the process of training achieve the expected student results, having good results and awaiting official certification.

The experience is developed in the subject Business Electronic (BE) which corresponds to the VIII semester of the Study Plan.

The objective is to design a feedback model of the problems solved by students in the Business Electronic course, in order to carry out an analysis of the recorded incidences, make decisions and take actions for improvement, which also allow the teacher to make decisions for continuous improvement in the teaching-learning process by solving problems by applying Problem Based Learning.

The research is descriptive, the methodology is based on the phases of the scientific method for problem solving [7], which states that the student has access to develop skills, and thus we would say greater knowledge of the problem area, procedures, attitudes and values.

As a result of working in groups and applying the PBL to the qualification of each of the problems posed, the students receive the Feedback Report of Evaluation of Deliverables together with the qualification matrix of the Criteria and their indicators involved to be analysed, discussed, reach conclusions on what to improve and determine their strategies to improve in the teaching-learning process. In addition to the statistics of the incidences derived from this.

The conclusion is that by applying PBL and evaluating the criteria and their indicators, together with a record of the incidents and the preparation of a Feedback Report for the students, which is sent to them as soon as possible, they help to correct mistakes or errors, increase their knowledge, strengthen the development of competences and soft skills, their attitudes, make their own judgements and assess the results of their work.

The article is organized in the Section II of Related Works - Context of the Experience, in Section III the Methodology is treated, in Section IV the Design of the Feedback Model is shown, in Section V the Method of Work is shown, in Section VI the Results of the work are shown, in Section VII the Discussion that is made is touched, that they worked or investigated others and of the result that is obtained, in Section VIII the conclusions of the work are shown, in Section IX the future works that can be carried out are shown.

II. Related Work – Context of the Experience

A. *Current Situation*

From 2019 we apply Formative Research adjacent to Active Didactic Strategies in the development of courses in the EPIS Given that the teaching learning process and competency based, training require feedback on the activities carried out by the students, this was done in a global way, not fulfilling the need to achieve significant and quality learning.

Thus, the evaluation of work, of each Deliverable Report, is based on the predefined criteria contemplated in the evaluation rubric, and thus a detailed, fair, real qualification, free as far as possible of subjectivities. This reflects the evolution and monitoring to be carried out on the group of students through adequate feedback. Thus, the problem is defined and will be considered in order to find a solution to the present work.

B. *Formative Research*

Formative research feeds research by training future researchers as stated by [8], that according to the professional training based on competences the student is at the centre of the process [9], with the teacher assuming the role of advisor with the rigor and demand in the development of the research.

In this sense, it is reiterated [10] that the mission of the university is to pay great attention to Formative Research, the actors directly involved are teachers and students, as it should be built throughout the process and thus deepen scientific research in the training of students for work performance in society.

There is experience in the EPIS of conducting Formative Research in courses such as Writing Articles and Research Reports (RAII), Research Methods and Writing, (MIR), Database, Business Electronic; the results are reflected in the Formative Research Report.

Applying active learning strategies; such as PBL; being participatory, whose focus is the student, generates methodological mechanisms to integrate it into the teaching-learning process [11], and develop research competences in the undergraduate whose role of Formative Inquiry is analysed by [12]. Together with the development of soft skills validated and evidenced by [13], it also contributes to developing the competences of the course.

C. *Student Competence Development*

All part of the educational model of the institution, in this case it is based on competences, which together with the Active Didactic Strategies as PBL make the competences to be developed appropriately by the students as referred by [14], we would say applied in different contexts.

Competence implies developing knowledge of an area of knowledge, skills and attitudes as referred to by [15]. If a methodology is also followed, as proposed by [16], the development of competences is demonstrated by solving problems of reality, discovering deficiencies or shortcomings, correcting weaknesses in the teaching learning process of engineering students supported by teachers. As well as valuing the achievements and learning from the experiences when carrying out research [17] by the students.

D. *Problem based Learning*

The training of students demands continuous updating so that when they graduate from the university they are able to work in different global contexts, to be part of multi- and trans disciplinary teams to solve problems of reality.

An alternative is to apply PBL, where the student is the centre of attention of the strategy [18], which in the area of engineering is used by offering interaction between teacher and student, and between students when working in groups of four to eight members. Working with PBL is based on the approach of a method such as that of [19] and [20] with steps to carry out the activities to be evaluated such as: reports, presentations, assessment of individual and team work as well as collaborative work, self-assessment, co-assessment, making judgements, critical thinking.

E. *Effective Evaluation*

If we start from the grade achieved in the resolution of a problem, task or work, this means approval, but it does not allow the student to know what his weaknesses and deficiencies are, which areas are weak in the subject treated, which concepts, methodologies, techniques, methods, the use of the appropriate tools, understanding his limitations, and what knowledge he should develop and deepen by researching in addition to what he has received in the class sessions.

In this regard [21] suggests that in order to have an effective evaluation, it must be:

- Valid; measuring what it should actually measure.
- Reliable; it must be consistent and fair.
- Transparent; that there are no tricks, traps or surprises.

The course contemplates and takes care that there is alignment between:

- Student Outcomes (SO).
- The defined contents to be covered in the course development.
- The learning activities in the theory and laboratory sessions to be carried out by the students.
- The evaluation method.
- The qualification tool that involves the general competences of the professional profile of the degree course, the course competences, the evaluation criteria

and their indicators, the qualification scales of the indicators.

F. Indicator based Evaluation of Learning

The assessment process generates information that when analysed, interpreted and then communicated generates knowledge with a value for action.

The assessment of learning should be a comprehensive and continuous process as discussed by [15] who considers basic questions such as What, How, When, Who to assess, the answers to which determine the type of assessment to be used, taking into account the context in which it takes place. This is also shared by [22] who highlights the finding of feedback as a formative act.

Thus, in the EPIS, professors apply Active Didactic Strategies that make it possible to evaluate student results in accordance with the ABET accreditation model, which works on the basis of criteria already defined for the courses of the Curricular Plan.

The Faculty of Production and Services Engineering (FIPS), to which the EPIS belongs, determined that the rubric should be used as an instrument, therefore in the present work the rubric [23, 24, 25, 26] was used to evaluate the criteria and its indicators related to student results, the achievement of increasing knowledge in the problem area, the development of competences, attitudes; for this purpose, rating scales are used according to the nature of the indicator to be rated in an objective and consistent manner.

It is through this assessment instrument that students are informed, in advance at the beginning of the course, which criteria are involved and must be fulfilled in order to reach the achievement levels of the Student Outcomes.

In the present work the BE course is taken as a case study applying the PBL in the laboratory sessions, taking as a result the evaluation of each Deliverable Report and recording and elaborating the Feedback Report addressed to the students.

G. Feedback in Learning

We have the contribution of [27] in the research of the literature review on feedback for learning. Thus it is considered the most salient as being:

a) Concept; citing Carless and Boud (2018) ... "deliver a definition of feedback from an understanding of a process through which learners use information to improve their work or learning strategies, this process being learner-driven in decision making to generate change (Dawson et al., 2019)".

b) Conceptual moments for Feedback.

- Feedback as a product; it is given by the teacher as the sole agent, it follows the idea of the corrective notion.
- Feedback as a dialogical act; the communication between teachers and students after the evaluation of a task based on criteria.
- Feedback as sustainable action; refers to interaction and dialogue as a support to the student in the tasks he/she performs; it is recursive in nature, it is delivered through cycles giving the opportunity to correct erroneous knowledge, leading to improvement, supported by clear criteria.

c) Feedback for Learning Model; providing feedback requires practice, through cycles in a sustainable way that allows learning beyond just completing the task. It is shown in Fig. 1.

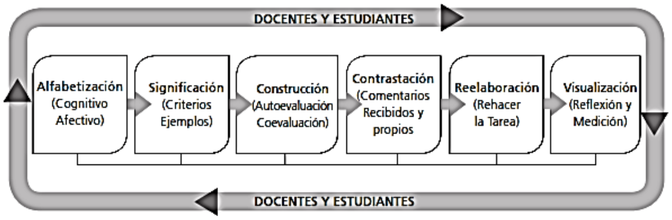

DOCENTES Y ESTUDIANTES
Alfabetización (Cognitivo Afectivo)
Significación (Criterios Ejemplos)
Construcción (Autoevaluación Coevaluación)
Contrastación (Comentarios Recibidos y propios
Reelaboración (Rehacer la Tarea)
Visualización (Reflexión y Medición)
DOCENTES Y ESTUDIANTES

Fig. 1. Feedback Model for Learning. Source: Quezada [27].

The definition of Feedback [28] is … Feedback is a process that helps to provide information about people's competencies, about what they know, what they do and how they act. Feedback allows us to describe how people think, feel and act in their environment and therefore allows us to know how it is performing and how it can be improved in the future."

In [29] citing Wiggins (2012) states characteristics that apply to the professor with respect to feedback such as:

a) Objective; the information given to the student should be related to the task being requested and conducive to learning.

b) Constructive; to consider the positive aspects, provide guidance, suggestions, how to overcome weaknesses, correct faults or mistakes.

c) Understandable; provide timely and detailed information on how to improve their learning.

d) Timely; the student receives feedback in time to improve their performance; or as we would say, to analyse and determine their strategies to improve their learning.

In educational evaluation, feedback is an essential component and is of great importance, as [30] points out that it is information that the student receives from each Deliverable Report of the BE course laboratory sessions about his or her performance. Having the following benefits it brings to providing the importance of feedback, such as:

- It clarifies the expected performance; the learning outcomes; based in our case on the analysis of the criteria and their indicators.
- It promotes dialogue between teacher and students, providing information about what has been achieved and allowing students to ask questions and thus improve their performance.
- Facilitates self-reflection; by instructing them to analyse the results of the scoring of the indicators of each criterion and thus address those where improvement is needed to achieve better results.
- Increases student motivation and self-esteem; students are motivated if through feedback they are shown the

positive aspects of their assessment and are told the level of marks they can achieve; they are also shown which points or aspects they need to improve and that they can make further progress, thus reinforcing their self-esteem.

The need arises for a procedure that provides students with a better feedback alternative in their course, compared to what is done in other courses in the curricula, and this is how the proposal of a simple and applicable feedback model arises, which has been maturing by applying the evaluation by indicators that make up a certain criterion.

III. Methodology

The methodological design is quasi-experimental, not employing a control group. The applied research is carried out to solve the problem of feedback through the Deliverable Evaluation Feedback Report. The research developed uses the feedback model generated, which involves the rubric instrument, the scoring tool for the indicators of each criterion of each defined competence of the subject.

IV. Design of the Feedback Model

The feedback model for learning should be delineated in dialogue, with the student gathering information from various sources about their performance and translating it into strategies for further improvement in their training.

The PBL uses a set of problems taken or adapted or posed by the professor, to be developed and solved by a group of students; for the course Business Electronic (BE) which is taken as a case for the present work.

Fig. 2 shows the Model developed and applied in semesters 2020 B and 2021 B in the BE course.

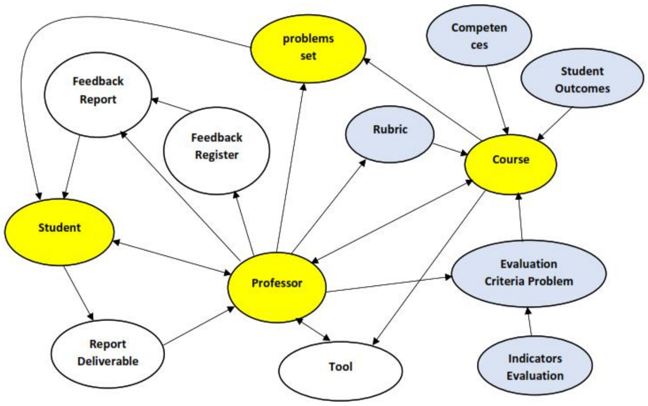


Fig. 2. Feedback Model. Source: Own Elaboration.

A. Reality of the Problem

Most of the evaluations carried out by professors teachers do not contain a documented record of the incidents, observations or suggestions made to the student or working group in the resolution of a problem that helps them to continuously improve in the development of the competences of the course involved.

B. Elements of the Model

The Feedback model includes the following elements:

1) Teacher
2) Student
3) Course
4) Competences
5) Student Outcomes to be assessed - criteria for SO
6) Evaluation Rubric - with criteria of the competences
7) Problem Evaluation Criteria
8) Evaluation Indicators of the Problem Evaluation Criteria
9) Problem Set
10) Qualification Tool
11) Deliverable Report
12) Deliverable Feedback Report Record
13) Deliverable Feedback Report.

The functions of each element of the model are detailed below:

- Professor
- Generates and updates the problem set
- Performs the grading of each Deliverable Report for each problem.
- Records the ratings in the Deliverable Feedback Report Log.
- Issues the Deliverable Feedback Report.

14) Student

- Form the working group for the academic semester.
- Work as a group to solve the proposed problems.
- Analyse, discuss the Deliverable Feedback Report, and make agreements for continuous improvement in the development of the course.

15) Course

- Contain the corresponding data and information for the development of the semester.

Involves a:

- Course competencies
- Student outcomes to be assessed - SO criteria
- Evaluation Rubric - with SO criteria
- Problem Evaluation Criteria
- Indicators of the problem's Evaluation Criteria.

16) Competences

- Contain the achievements to be attained by the students according to the general competences of the Professional Career Profile.

17) Student outcomes to be assessed - criteria of SO

- Contain the Student Outcomes expected to be achieved in the student's education in each course of the Curriculum.

*18)*Evaluation Rubric - with competency criteria

- Contain the evaluation criteria for each competency for the qualification of the Deliverable Report for each problem posed in the course.

*19)*Evaluation Criteria of the problem

- Generate the frame of reference for professors to evaluate the academic performance of students.
- Define the knowledge that students must acquire, what they have to learn to perform or create with such knowledge.
- They are those defined to be evaluated for each competence for the qualification of the Deliverable Report of each problem posed in the course.

*20)*Evaluation Indicators of the Evaluation Criteria of the problem

- Contain the indicators that are the most specific defined within a criterion and that will be evaluated from each Evaluation Criterion for the qualification of the Deliverable Report of each problem posed in the course.

*21)*Problem Set

- Contains the set of problems to be developed by the students during the semester.

*22)*Qualification Tool

Based on the Electronic Spreadsheet - EXCEL.

Involves:

- Course
- Course Competences
- Problem Evaluation Criteria
- Indicators of the Problem Evaluation Criteria.
- Indicator qualification scales.

Its responsibilities are to:

- Storing the marking data for the criteria indicators for each competency, for each proposed problem to be worked on.
- Calculate and record the continuous evaluation and examination marks for each semester evaluation period of the course.
- Calculate and record the marks for each Student Outcomes Criterion involved.

*23)*Deliverable Report

- Contain the development of the constituent parts of the Deliverable Report according to a predefined structure.

*24)*Deliverables Feedback Report Record

- Contain the observations, suggestions, contributions derived from the qualification of each indicator of each criterion of the Deliverable Report.

*25)*Deliverable Feedback Report

- Contain the observations, suggestions and inputs from the Feedback Report Register and from the qualification of the indicators of each criterion.

C. The Project

The design of the project contemplated the aspects considered in the professor's evaluation of the Deliverables Reports containing the criteria and their indicators, contemplating the application of the indicator-based evaluation system of [31]. And of the application of the system taking the results shown by [32] of the experience of evaluation based on indicators.

The developed project has the following characteristics:

- It contains problems of reality organisations or adaptations or case studies.
- The work team is made up of four or three students, taking into account the availability of time and timetables of its members, and also taking into account the other courses of study and personal interrelation, which provides very good results.
- The problems are developed during the academic semester, preparing the Deliverable Report for each one.
- Carry out the analysis of the Feedback Report and take actions for the continuous improvement of the Deliverable Reports to be elaborated in the following sessions.

D. Incidences

The table of incidences (observation) has been determined according to the evolution of the qualification of the indicators in the work done in the semesters involved, and what each teacher has found and how he/she has applied in the recording of incidences. Table I shows the incidences.

TABLE I. INCIDENCES

Codig	Description
1	Absent
2	Very incipient
3	Insufficient
4	Incomplete
5	No further explanation
6	No guard coherence
7	Stake out
8	Are not such
9	Inconsistent
10	Improve to reach excellent
11	To draft adequately

Source: Own Elaboration.

E. Follow-Up

Sessions are held to deliver the Deliverable Report during the semester, which contains the result of the problem, the presentation, the evaluation with the corresponding qualification of the indicators of each criterion, and at the same time the Feedback Report is drawn up. This is distributed to the members of the team involved so that in a joint session they can analyse the evaluation carried out by the teacher, and determine the shortcomings and weaknesses which must be corrected, taking into account the suggestions and opinions; and on this basis propose their strategies for continuous improvement in the development of the problems and therefore of the following deliverable reports.

As a principle of PBL, the teacher plays the role of a coach by providing suggestions, guidelines, pointing out shortcomings, giving guidance and comments to the students.

The qualification of the Deliverable Reports is carried out on the basis of the work done by [32] and at the same time, in parallel, the incidences of the indicators of each criterion are recorded in the Feedback Report of each group of students' work.

V. METHOD OF WORK

A. Conceptual Design

Having defined the Student Outcomes for each course of the Curriculum Plan, the teacher must propose the products, behaviours, actions or others that will be requested to the group of students in order to know the level of achievement.

The purpose is for the student to investigate, as referred to in [33] by carrying out research activities and using PBL to solve problems in order to discover more knowledge related to the problem area, reinforce other concepts, develop competences [33] soft and procedural skills, make their own judgements, and evaluate the results achieved.

The research for the development of the work is based on the stages of problem solving, of the scientific method, which associated with the active didactic strategy of PBL, the group of students develop each problem posed and elaborate the Deliverable Report following the stages of the project proposed by [32] in order for the professor to elaborate the Feedback Report.

In order to obtain the data on the students' assessment of the Feedback Model, the survey technique and its instrument, the questionnaire, are used, and then the results are systematised in order to analyse them and draw conclusions that can be used to make decisions regarding the continuous improvement of the teaching-learning process of the course.

Methodologies, Techniques and Instruments to be used

A Methodologies, Techniques and Instruments

- Problem Based Learning Methodology - PBL.
  - Problem solving, in sessions determined in the development of the course.
- Survey Technique
  - Questionnaire - Questionnaire of Assessment of the Feedback Model for the Application of Problem-Based Learning.

B Evidence

- Digital files, of the Reports of Deliverables of the laboratory work.
- Digital files, of the records of Incidents or Observations for the Deliverables Evaluation Feedback Report.
- Digital files of the Feedback Reports.
- Moodle, Virtual Classroom as a repository of the course work.

Evaluation Instruments

- Deliverable Report Qualification Tool.
- Evaluation rubrics.
  - Evaluation of Competences and Student Outcomes.

B. Participants

The BE course in the EPIS is taken as a case study, which is developed in the eighth semester (4th year), with five hours per week, three theoretical and two laboratory hours, in 17 weeks, the theory and laboratory practices were carried out by two professors. The students participated in the elaboration of the Deliverable Report, taking as a case study the semester 2020 B with 10 subgroups and a total of 35 students; the semester 2021 B with 15 subgroups and a total of 57 students.

C. Data Analysis Technique

From the registration of the incidences by the qualification of the indicators of each criterion of the developed problems, the data were systematised in the electronic spreadsheet EXCEL, visualising and analysing the results achieving frequencies, averages, tables, graphs.

D. Instruments

The following instruments are used:

- Template for the elaboration of the Feedback Report.
- Evaluation rubrics.
- Deliverable Report Qualification Tool.
- Feedback Model Valuation Questionnaire.

E. Techniques

Evaluation techniques are used:

- The rubric.
- The qualification scale.
- Survey.

F. Deliverables

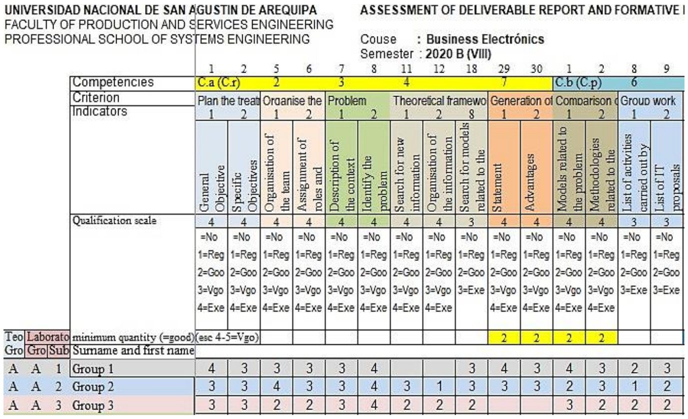
It was established that the Feedback Report of the evaluation of the deliverables will be delivered to each working subgroup and stored in the repository of the virtual classroom.

The structure of the Feedback Report contains two parts:

- Feedback Report of each Deliverable Report
- Evaluation of Deliverable Report.

VI. RESULTS

The results of the Evaluation of the Deliverable Report are shown below; with a defined and used structure, based on the rubric defined for the BE course in semester 2020 B. Fig. 3 shows a sample of the course competences that are related to the general competences of the course, the criteria, the indicators of the criteria with the qualification scale determined for each one of them, and the qualification of the proposed problem of the working groups.



UNIVERSIDAD NACIONAL DE SAN AGUSTIN DE AREQUIPA
FACULTY OF PRODUCTION AND SERVICES ENGINEERING
PROFESSIONAL SCHOOL OF SYSTEMS ENGINEERING

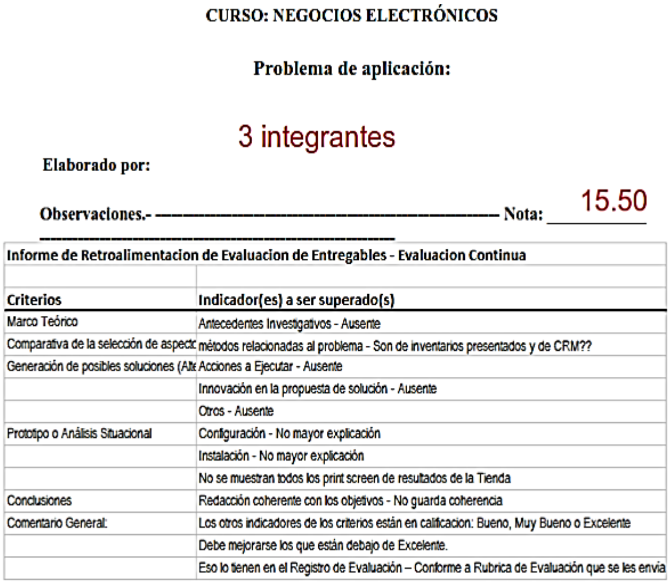
ASSESSMENT OF DELIVERABLE REPORT AND FORMATIVE I

Couse : Business Electrónics
Semester : 2020 B (VIII)

				1	2	5	6	7	8	11	12	18	29	30	1	2	8	9
			Competencies	C.a (C.r)		2		3		4			7		C.b (C.p)		6	
			Criterion	Plan the treat		Organise the		Problem		Theoretical framewo			Generation of		Comparison (		Group work	
			Indicators	1	2	1	2	1	2	1	2	8	1	2	1	2	1	2
				General Objective	Specific Objectives	Organisation of the team	Assignment of roles and	Description of the context	Identify the problem	Search for new information	Organisation of the information	Search for models related to the	Statement	Advantages	Models related to the problem	Methodologies related to the	List of activities carried out by	List of IT proposals
			Qualification scale	4	4	4	4	4	4	4	4	3	4	4	4	4	3	3
				=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Vgo 4=Exe	=No 1=Reg 2=Goo 3=Exe	=No 1=Reg 2=Goo 3=Exe
Teo	Laborato		minimum quantity (=good)	(esc 4-5=Vgo)									2	2	2	2		
Gro	Gro	Sub	Surname and first name															
A	A	1	Group 1	4	3	3	3	3	4			3	4	3	4	3	2	3
A	A	2	Group 2	3	3	4	3	3	4	3	1	3	3	3	2	3	1	2
A	A	3	Group 3	3	3	2	2	3	4	2	2	2			3	2	2	2

Fig. 3. Report of Assessment of Deliverable Report 2020 B with Qualification Tool. Source: Own Elaboration.

Then the results of the evaluation of the Deliverable Report are shown; which are reflected in the Feedback Report implemented in semester 2020 B showing the criteria and indicators with observation of the incidences found as the professor includes them always based on the defined rubric of the BE course. In Fig. 4 there is a sample, where the Feedback Report is included in the cover of the Deliverable Report for its registration and storage and its qualification resulting from applying the qualification tool of the proposed problem of the working group.



CURSO: NEGOCIOS ELECTRÓNICOS

Problema de aplicación:

3 integrantes

Elaborado por:

Observaciones.- -- **Nota:** 15.50

Informe de Retroalimentacion de Evaluacion de Entregables - Evaluacion Continua

Criterios	**Indicador(es) a ser superado(s)**
Marco Teórico	Antecedentes Investigativos - Ausente
Comparativa de la selección de aspectc	métodos relacionadas al problema - Son de inventarios presentados y de CRM??
Generación de posibles soluciones (Alte	Acciones a Ejecutar - Ausente
	Innovación en la propuesta de solución - Ausente
	Otros - Ausente
Prototipo o Análisis Situacional	Configuración - No mayor explicación
	Instalación - No mayor explicación
	No se muestran todos los print screen de resultados de la Tienda
Conclusiones	Redacción coherente con los objetivos - No guarda coherencia
Comentario General:	Los otros indicadores de los criterios están en calificacion: Bueno, Muy Bueno o Excelente
	Debe mejorarse los que están debajo de Excelente.
	Eso lo tienen en el Registro de Evaluación – Conforme a Rubrica de Evaluación que se les envia

Fig. 4. Deliverable Report Including Rating Observations - Semester 2020 B. Source: Own Elaboration.

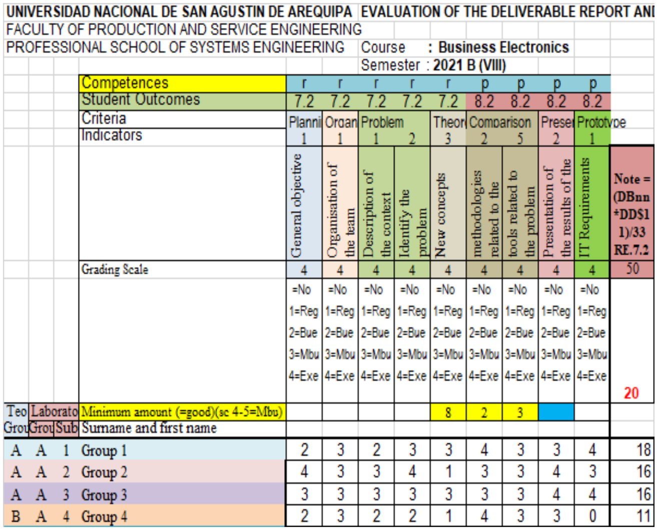
According to the analysis and conclusions reached from the development of semester 2020 B, the relevant modifications and inclusions were made to improve the grading tool, as well as the Feedback Report to make it more practical in its registration, by creating a coding of the criteria and indicators that serves for the systematisation of the same; and that serves as a better communication between the professor and the student, being clearer and more understandable, as shown in Fig. 5 and Fig. 6.

The results of the Evaluation of the Deliverable Report are shown below; with a structure determined and used, based on the defined rubric of the BE course of semester 2021 B. In Fig. 5 there is a sample, where the Competences of the course related to those of the degree course are shown, the inclusion of the Student Results to be achieved, the Criteria and the Indicators with the Qualification Scale determined for each one of them, and their qualification carried out of the problem proposed to the work groups; which at the end is reflected in the marks obtained for each student result, as well as the marks for the resolution of each problem posed.

Fig. 6 shows the results of the evaluation of the Deliverable Report; with the observations of the Incidences determined or found of the criteria and indicators involved and are reflected in the Feedback Report implemented in semester 2021 B, which in its design is more explicit and improves the communication between the professor and the students.

The Feedback Report that is sent to the members of each of the subgroups consists of the content of Fig. 6 Deliverable Report including incidents of the Feedback Report and of the Fig. 5 Evaluation of the Deliverable Report.

As a further result, Tables II and III reflect the performance results according to the level of ABET accreditation performance rating for the EPIS courses, which are at the satisfactory and outstanding level, showing that the Student Outcomes are achieved, as well as the development of the Competences and the Feedback Report plays an important role in achieving this.

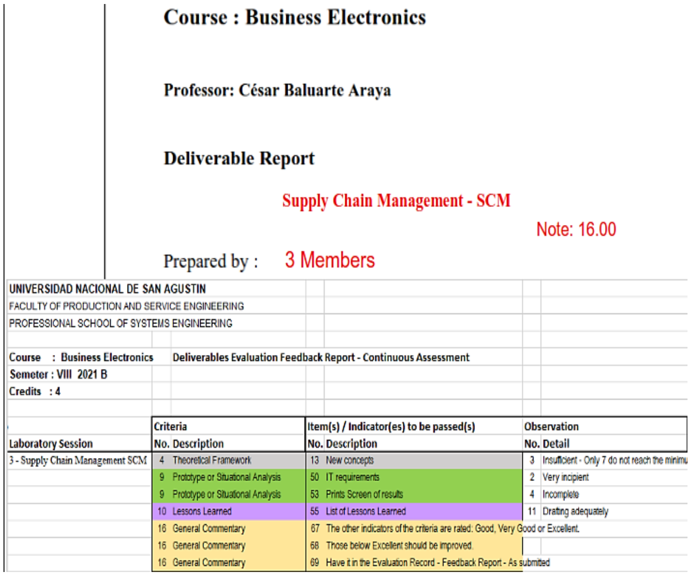


UNIVERSIDAD NACIONAL DE SAN AGUSTIN DE AREQUIPA — EVALUATION OF THE DELIVERABLE REPORT ANI
FACULTY OF PRODUCTION AND SERVICE ENGINEERING
PROFESSIONAL SCHOOL OF SYSTEMS ENGINEERING

Course : Business Electronics
Semester : 2021 B (VIII)

			Competences	r	r	r	r	r	p	p	p	p	
			Student Outcomes	7.2	7.2	7.2	7.2	7.2	8.2	8.2	8.2	8.2	
			Criteria	Planni	Organ	Problem		Theor	Comparison		Preser	Prototype	
			Indicators	1	1	1	2	3	2	5	2	1	
				General objective	Organisation of the team	Description of the context	Identify the problem	New concepts	methodologies related to the	tools related to the problem	Presentation of the results of the	IT Requirements	Note = (DBnn *DD$1 1)/33 RE.7.2
			Grading Scale	4	4	4	4	4	4	4	4	4	50
				=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	=No 1=Reg 2=Bue 3=Mbu 4=Exe	20
Teo	Laborato		Minimum amount (=good)(sc 4-5=Mbu)					8	2	3			
Grou	Grou	Sub	Surname and first name										
A	A	1	Group 1	2	3	2	3	3	4	3	3	4	18
A	A	2	Group 2	4	3	3	4	1	3	3	4	3	16
A	A	3	Group 3	3	3	3	3	3	3	3	4	4	16
B	A	4	Group 4	2	3	2	2	1	4	3	3	0	11

Fig. 5. Report of Assessment of Deliverable Report 2021 B with Qualification Tool. Source: Own Elaboration.



Course : Business Electronics

Professor: César Baluarte Araya

Deliverable Report

Supply Chain Management - SCM

Note: 16.00

Prepared by : 3 Members

UNIVERSIDAD NACIONAL DE SAN AGUSTIN
FACULTY OF PRODUCTION AND SERVICE ENGINEERING
PROFESSIONAL SCHOOL OF SYSTEMS ENGINEERING

Course : Business Electronics — Deliverables Evaluation Feedback Report - Continuous Assessment
Semeter : VIII 2021 B
Credits : 4

	Criteria		Item(s) / Indicator(es) to be passed(s)		Observation	
Laboratory Session	No.	Description	No.	Description	No.	Detail
3 - Supply Chain Management SCM	4	Theoretical Framework	13	New concepts	3	Insuficient - Only 7 do not reach the minimu
	9	Prototype or Situational Analysis	50	IT requirements	2	Very incipient
	9	Prototype or Situational Analysis	53	Prints Screen of results	4	Incomplete
	10	Lessons Learned	55	List of Lessons Learned	11	Drafting adequately
	16	General Commentary	67	The other indicators of the criteria are rated: Good, Very Good or Excellent.		
	16	General Commentary	68	Those below Excellent should be improved.		
	16	General Commentary	69	Have it in the Evaluation Record - Feedback Report - As submitted		

Fig. 6. Deliverable Report Including Incidents of the Feedback Report - Semester 2021 B. Source: Own Elaboration.

TABLE II. PERFORMANCE REPORT OF THE BUSINESS ELECTRONICS 2020 B COURSE - ABET

	Number Students	**%**	**%**	**Performance**
Grade Obtained	**Global**	**Global**	**Evaluated**	**Qualificated**
AB Abandonment	0	0	0,00	Abandonment
From 1 to 7	0	0	0.00	Unsatisfactory
From 8 to 10	0	0	0.00	In Progress
From 11 to 14	3	8.5714	8.57	Satisfactory
From 15 to 20	32	91.4286	91.43	Outstanding
	35	100.0000	100.00	

Source: Own Elaboration.

TABLE III. PERFORMANCE REPORT OF THE BUSINESS ELECTRONICS 2021 B COURSE - ABET

	Number Students	**%**	**%**	**Performance**
Grade Obtained	**Global**	**Global**	**Evaluated**	**Qualificated**
AB Abandonment	1	1.7544	1.75	Abandonment
From 1 to 7	0	0	0.00	Unsatisfactory
From 8 to 10	0	0	0.00	In Progress
From 11 to 14	15	26.3158	26.32	Satisfactory
From 15 to 20	41	71.9298	71.93	Outstanding
	57	100.0000	100.00	

Source: Own Elaboration.

Fig. 7 shows the progress of group three in the development of the problems set during the 2020 B academic semester.

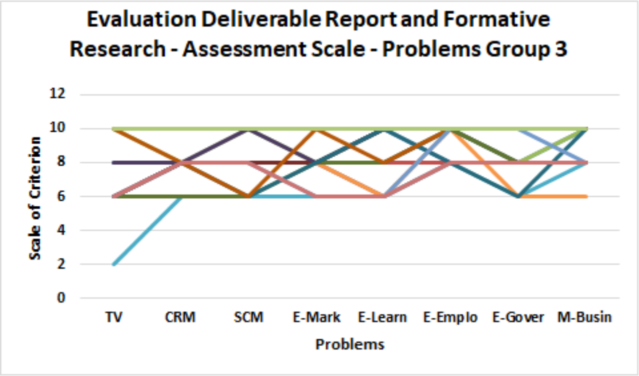


Fig. 7. EDRFR – AS – PG3 - Evaluation Deliverable Report and Formative Research-Assessment Scale-Problems Group 3, Semester 2020 B. Source: Own Elaboration.

Fig. 8 also shows how in one way or another the progress in this case of group 21, in the development of the problems that were posed for the academic semester 2021 B.

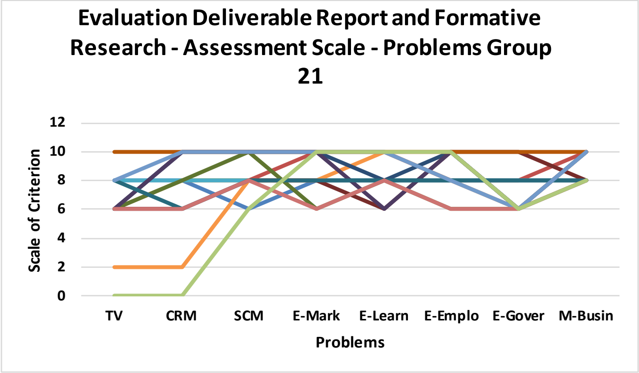


Fig. 8. EDRFR – AS – PG21 - Evaluation Deliverable Report and Formative Research-Assessment Scale-Problems Group 3, Semester 2021 B. Source: Own Elaboration.

VII. DISCUSSION

There are works on the application of didactic strategies such as that of [34] emphasising in his book the relevant explanations to achieve the objectives set out in each context where the teaching-learning process takes place.

The proposal of [35] is to go beyond a simple evaluation and use didactic teaching-learning strategies applied to the context of reality that contribute to student learning.

In the research, the application of PBL in the work of [32] has been achieved by having, at the moment of the qualification of the indicators of each criterion, the possibility of determining the incidences and observations to which it gives rise and registering them in the corresponding medium..

We have the results of the evaluation grades of the last four years, which show how the qualification based on indicators allow a fairer and more objective evaluation of each student's Deliverable Report, when applying PBL from 2019, which are shown in Table IV regarding the Mobile Business problem, being on average at an outstanding level according to the scale determined for the measurement of student performance (the minimum is 15) in the EPIS courses.

TABLE IV. GRADE POINT AVERAGE

Course	**Grade Point Average**			
	2018	***2019***	***2020***	***2021***
Business Electronics Theme: Mobile Business	15.17	16.15	17.4	15.10

Source: Own Elaboration.

Also [36] presents some general principles of effective feedback, and how, as he calls it, the Decalogue, helps students to improve and supports their self-esteem, which we consider a good contribution to be contemplated by the professor. Also taking into account what has been discussed by [37] regarding the principles of effective feedback.

VIII. CONCLUSION

The following conclusions have been reached:

The objectives of the Feedback Report have been achieved by reaching a better communication between the professor and the student so that they can formulate their strategies and take the relevant actions for continuous improvement.

The Student Outcomes determined for the course are achieved at an outstanding level; and students achieve through feedback a better development of their competences as well as those of the course, and thus also achieve the competences of the Professional School's syllabus.

Problem Based Learning as an active methodology and the Feedback Report allow; to increase motivation for autonomous learning, to increase knowledge on topics of the problems of the area, to apply the knowledge achieved, motivation for teamwork, to complement the development of soft and procedural skills, progressively improving their academic performance.

Through the evaluation of the Deliverable Report, students obtain an evaluation according to the qualification of each indicator in the criteria involved in each problem, which is more objective, fair, real and with less subjectivity on the part of the qualifier; students recognising that the feedback provided in the Feedback Report helps them to improve and perform better in teamwork.

Providing students with feedback in the Deliverable Reports on the reviews, qualifications and observations made by the professor, allows them to continuously improve in the work group.

IX. FUTURE WORK

To carry out comparative research on the application of PBL and others such as PrBL (Project Based Learning) in the courses of the Curriculum Plan using the evaluation by indicators and its Feedback Report.

The systematisation of the criteria and their indicators as well as the result of the work of [32] will allow the subsequent development of an evaluation and feedback system.

ACKNOWLEDGMENT

Our thanks to the Universidad Nacional de San Agustín de Arequipa for allowing and supporting the development of Formative Research and the application of the proposals for continuous improvement based on Active Didactic Strategies with the favourable results being of benefit to the students and the teaching-learning process.

REFERENCES

[1] Universidad Nacional de San Agustín de Arequipa. http://www.unsa.edu.pe.

[2] Subdirección de Currículum y Evaluación, Dirección de Desarrollo Académico, Vicerrectoría Académica de Pregrado, Universidad Tecnológica de Chile INACAP. (2017), pp. 21-24. *Manual de Estrategias Didácticas: Orientaciones para su selección*. Santiago, Chile: Ediciones INACAP.

[3] R. Rodriguez Cruz, Compendio de estrategias bajo el enfoque por competencias, Instituto Tecnológico de Sonora ITESCA, México, 2007, pp. 26-29. http://www.itesca.edu.mx/documentos/desarrollo_academico/compendio_de_estrategias_didacticas.pdf.

[4] C Ma. Cristina Sánchez Martínez, M. Aguilar Venegas, J.L. Martínez Durán and J.L. Sánchez Ríos, Estrategias didácticas en entornos de aprendizaje enriquecidos con tecnología (antes del Covid-19), UNIVERSIDAD AUTÓNOMA METROPOLITANA-XOCHIMILCO, México, 2020, pp. 11-15.

[5] Escuela Profesional de Ingeniería de Sistemas. http://www.episunsa.edu.pe.

[6] ABET. Why ABET Accreditation Matters. https://www.abet.org/accreditation/what-is-accreditation/why-abet-accreditation-matters/. Ultimo acceso junio 2021. https://www.abet.org/assessment/.

[7] RG. GARZA-RIVERA, El rol de la física en la formación del ingeniero. *Ingenierías*, 2001, vol. IV, No. 13, pp. 48-54.

[8] R.A. Patiño, Z.A. Melgarejo, and G.M. Valero, (2019). Percepción de los egresados contables sobre la investigación formativa. Revista *Activos, 16*(30), 101-125. DOI: https://doi.org/10.15332/25005278.5062, Colombia.

[9] F. Medina-Rojas, J.M. Nuñez-Santa, I.I. Sánchez-Medina, and J.M. Cabrera-Medina, Implementación del ABP, PBL y método SCRUM en cursos académicos para desarrollar sistemas informáticos enfocados en fortalecer la región, Revista Educación en Ingeniería, 12(24), pp. 52-57, Julio, 2017, Bogotá. ISSN 1900-8260. DOI: http://dx.doi.org/10.26507/rei.v12n24.758.

[10] B. Restrepo Gómez, INVESTIGACIÓN FORMATIVA E INVESTIGACIÓN PRODUCTIVA DE CONOCIMIENTO EN LA UNIVERSIDAD, Nómadas (Col), núm. 18, mayo, 2003, pp. 195-202, ISSN: 0121-7550, Universidad Central, Bogotá, Colombia.

[11] C. Mejía Murillo, Manual de Procesos de Investigación Formativa, Universidad Herminio Valdizán, Perú, 2016, pp. 7-9.

[12] A. Pinto Santos, and O. Cortés Peña, ¿Qué Piensan los Estudiantes Universitarios Frente a la Formación Investigativa?, *REDU. Revista de Docencia Universitaria*, 2017, *15*(2), 57-75.

[13] C. Baluarte Araya, E. Vidal Duarte, and E. Castro Gutierrez, Validación de las Habilidades Blandas en los cursos de la Currícula de la Escuela Profesional de Ingeniería de Sistemas-UNSA, 16th LACCEI International Multi-Conference for Engineering, Education, and Technology: "Innovation in Education and Inclusion", 19-21 July 2018, Lima, Perú. Digital Object Identifier (DOI): http://dx.doi.org/10.18687/LACCEI2018.1.1.97 ISBN: 978-0-9993443-1-6 ISSN: 2414-6390.

[14] E. Briones, and J. Vera, Aprendizaje Basado en Problemas (ABP): Percepción de carga de trabajo y satisfacción con la metodología, V Congreso Mundial de Estilos de Aprendizaje, Santander, España,2012. https://dialnet.unirioja.es/servlet/articulo?codigo=4640627.

[15] M.J. Acebedo, (2016). La evaluación del aprendizaje en la perspectiva de las competencias. Revista TEMAS, 3(11), pp. 203–226.

[16] F.H. Fernández, and J. Duarte, El aprendizaje basado en problemas como estrategia para el desarrollo de competencias específicas en estudiantes de ingeniería, Formación Universitaria, 6(5), 2013, pp. 29-38.

[17] C. Gamboa, "Apuntes sobre Investigación Formativa"; versión No. 2, Colombia, 2013. [On line]. Disponible: http://idead.ut.edu.co/Aplicativos/PortafoliosV2/Autoformacion/materiales/documentos/u2/Apuntes_sobre_investigacion_formativa.pdf.

[18] B. Restrepo Gómez, Aprendizaje basado en problemas (ABP): una innovación didáctica para la enseñanza universitaria, Educación y Educadores, vol. 8, 2005, pp. 9-19, Universidad de La Sabana, Cundinamarca, Colombia.

[19] P. Morales and V. Landa, "Aprendizaje Basado en Problemas". Theoria, Vol. 13, 2004, pp. 145-157. [On line]. Disponible: http://biblioteca.udgvirtual.udg.mx/jspui/handle/123456789/574.

[20] Dirección de Investigación y Desarrollo Educativo, Vicerrectoría Académica, Instituto Tecnológico y de Estudios Superiores de Monterrey. "El Aprendizaje Basado en Problemas como técnica didáctica", pp.14-18, [On line]. Disponible: http://www.sistema.itesm.mx/va/dide/inf-doc/estrategias.

[21] L. Jabif, Centro de Actualización en la Enseñanza Superior (CAES), Universidad ORT, Uruguay. https://caes.ort.edu.uy/herramientas-para-la-docencia/sugerencias-para-una-evaluacion-efectiva. Consulted 16 12 2020.

[22] C. Román, Sobre la retroalimentación o el feedback en la educación superior on line, Revista Virtual Universidad Católica del Norte, núm. 26, febrero-mayo, 2009, pp. 1-18, Fundación Universitaria Católica del Norte,Medellín, Colombia.

[23] Octaedro, Rúbricas para la evaluación de competencias, España, 2013, pp. 8-23.

[24] M.C. Sáiz Manzanares and A. Bol Arreba, Aprendizaje basado en la evaluación mediante rúbricas en educación superior, Elsevier, Suma Psicológica, SUMA PSICOL. 2014; 21(1**)**:28-35, España, 2014.

[25] NA Ortega Andrade, MA Romero Ramírez, and RME Guzmán Saldaña, Rubrica para evaluar la elaboración de un Proyecto de Investigación Basado en el Desarrollo de Competencias, Universidad Autónoma del estado de Hidalgo, México, 2014, vol.2, No. 4. https://www.uaeh.edu.mx/scige/boletin/icsa/n4/e6.html.

[26] Ma.C. Sanchez Martinez, M. Aguilar Benegas, J.L. Martinez Durán, and J.L. Sánchez Ríos, Estrategias didácticas en entornos de aprendizaje enriquecidos con tecnología (antes del covid-19), Universidad Autónoma Metropolitana-Xochimilco, México, 2020, pp.67-71.

[27] S. Quezada, and C. Salinas, Modelo de Retroalimentación para el Aprendizaje, Revista Mexicana de Investigación Educativa, RMIE, 2021, VOL. 26, NÚM. 88, PP. 225-251 (ISSN: 14056666 • ISSN-e: 25942271).

[28] P. Avila, La Importancia de la Retroalimentación en los Procesos de Evaluación, Universidad del Valle, 2009, México. http://www.universidadcies.com/wp-content/uploads/2017/06/Avila_retroalimentacion.pdf. Consulted 18 05 2022.

[29] S. Valdivia, En blanco & Negro, Revista sobre Docencia Universitaria, (2014) Vol. 5 N° 2, Pontificia Universidad Católica del Perú - PUCP, Perú. Consulted 31 03 2022.

[30] L. Jabif, Centro de Actualización en la Enseñanza Superior (CAES), Universidad ORT, Uruguay. https://caes.ort.edu.uy/herramientas-para-la-docencia/la-importancia-de-la-retroalimentacion. Consulted 31 05 2022.

[31] C. Baluarte-Araya, Proposal of an Assessment System based on Indicators to Problem Based Learning – IEEE Conference Publication, Published in: 2020 39th International Conference of the Chilean Computer Science Society (SCCC), 16-20 Nov. 2020, Coquimbo, Chile. DOI: 10.1109/SCCC51225.2020.9281203.

[32] C. Baluarte-Araya, E. Suarez-Lopez, and O. Ramirez-Valdez, Problem based Learning An Experience of Evaluation based on Indicators, Case of Electronic Business in Professional Career of Systems Engineering, (IJACSA) International Journal of Advanced Computer Science and Applications, Volume 12 Issue 9, 2021, pp. 581-592. DOI: 10.14569/IJACSA.2021.0120966.

[33] C. Baluarte-Araya and N. Bedregal-Alpaca, Influence of Problem Based Learning on the development of technical, methodological, participatory and personal competencies valuation by Engineering students, presented at VI IEEE World Engineering Education Conference (EDUNINE2022), Smart Distributed Conference in IBERO-AMERICA from March 13 to 16, 2022. In press.

[34] J. Flores, J. Avila, C. Rojas, F. Sáez, R. Acosta, and C. Diaz, Estrategias Didácticas para el aprendizaje significativo en contextos universitarios, Unidad de Investigación y Desarrollo Docente, Dirección de Docencia, Universidad de Concepción, Chile, 2017, pp. 10-14.

[35] B. Santamarina, La evaluación de los estudiantes en la Educación Superior: Mas allá de la Evaluación, Universidad de Valencia, Servicio de Formación Permanente, España, 2007, pp. 49-53.

[36] G. Contreras, Proyecto MECESUP, Pontificia Universidad Católica de Valparaiso, pp. 161-175, 2014, Chile.

[37] C. Castro, Evaluación y Retroalimentación para los aprendizajes, Instituto Profesional IACC, Universidad de Chile 2020. https://educacionsuperior.mineduc.cl/wp-content/uploads/sites/49/2020/04/6-Modelo-Evaluacion-y-retroalimentacion-aprendizajes.pdf. Consulted 01 04 2022.

Development of Discrepancy Evaluation Model based on Tat Twam Asi with TOPSIS Calculation

Dewa Gede Hendra Divayana[1]*, Agus Adiarta[2], P. Wayan Arta Suyasa[3]
Department of IT Education, Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia[1, 3]
Department of Electrical Education, Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia[2]

***Abstract*—This research had the main objective to provide information related to the innovation available in the form of an educational evaluation model that integrates the Discrepancy evaluation component, Tat Twam Asi concept, and TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method in the framework of determining the dominant indicators triggering the effectiveness of implementing blended learning in IT vocational schools. The approach of this research was development research by an R & D development model that focused on four stages, including a) research and field data collection, b) planning, c) design development, d) initial trial, and e) revisions to the results of the initial trial. There were 34 subjects involved in the trial design of the evaluation model in this research, including two education experts, two informatics experts, and 30 IT vocational teachers in Bali. The instruments used in data collection were in the form of questionnaires, interview guidelines, and photo documentation. The analysis technique for the data that had been collected used quantitative descriptive techniques that referred to percentage descriptive calculations. The results of this research were Tat Twam Asi-based Discrepancy evaluation model design which was integrated with TOPSIS calculations and had been classified as excellent according to the eleven-scale categorization table.**

Keywords—Discrepancy; evaluation model; tat twam asi; TOPSIS

I. Introduction

Nowadays blended learning has become a vital requirement in the learning process at IT vocational schools because of the demands for flexibility, convenience, speed, and transparency in the educational field as a result of the appearance of industrial revolution 4.0. The fact shows that the blended learning implementation in some IT vocational schools was not optimal. It is following the statement of Mozelius and Rydell [1], who stated that “there were still many cases that show that blended learning has not been implemented well in the learning process”. Even though evaluation activities were often carried out in the blended learning implementation, the recommendations given were not yet precise regarding the target, especially in determining the dominant indicators that trigger the level of blended learning effectiveness. Several evaluation models have been used by educational evaluators to evaluate the blended learning implementation, including CSE-UCLA [2], CIPP [3], and Formative-Summative [4]. However, among those models, an exact model has not yet been found in determining the dominant indicators that trigger the effectiveness of blended learning based on the weighting equation given by evaluators to the defining, installation, process, and product components. One innovation to overcome those problems was to use the Tat Twam Asi-based Discrepancy evaluation model with an accurate and systematic calculation process using the TOPSIS method so that a dominant indicator can be determined as a trigger for the blended learning effectiveness. The discrepancy model can show the evaluation components, including definition, installation, process, and product. The concept of Tat Twam Asi (a local wisdom concept in Bali that means I am you) adheres to the philosophy of equality which can be used in determining the weighting equation given by evaluators. The TOPSIS method can be used to determine dominant indicators based on the highest preference value of each evaluation indicator. From the innovation findings in overcoming those problems, the research problem was “How was the design of Tat Twam Asi based Discrepancy evaluation model by TOPSIS calculation in determining the dominant indicators triggering the effectiveness of blended learning implementation in vocational high school (case study in Bali province)?”

This research was motivated by the results of the following studies, including (1) research in 2017 conducted by Embi et al. [5] showed that there was a deep assessment using the Kirkpatrick model in evaluating the implementation of multimedia-based blended learning. Limitation of the Embi et al.’s research has not yet shown in detail the assessment indicators that were the priority determinant of the multimedia-based blended learning effectiveness; (2) research conducted in 2018 by Istanbul and Supriadi [6] showed the use of the CIPP model to evaluate the blended learning implementation that supports the learning process at Widyatama University. The limitation shown in Istanbul and Supriadi’s research was that priority indicators have not been shown to trigger the successful implementation of blended learning in the learning process; (3) research conducted in 2019 by Agustina and Mukhtaruddin [7] basically showed four evaluation components that were the same as the evaluation components used in this research, including defining (same with Context component), installation (same with Input component), process (same with Process component), and product (same with Product component). The four components serve as the basis for evaluating the blended learning implementation that was used to support the integrated English learning process. The limitation of Agustina and Mukhtaruddin’s research was that they have not been able to show the most dominant indicators as triggers for the effectiveness of blended learning implementation; (4) research conducted in 2019 by Ngala et al. [8] showed the use of the CIPP model to evaluate the implementation of distance education based on e-learning and

*Corresponding Author

blended learning. The limitation of Ngala et al.'s research was not yet showing the evaluation standards in detail and had not been able to show the dominant indicators that trigger the successful implementation of distance education; (5) research conducted in 2019 by Siswadi et al. [9] showed the limitations of the CIPP model, especially on aspects in the context component and the input component used in evaluating national standards of nursing education. Besides, in research of Siswadi et al. also has not shown any aspects or indicators that trigger the effectiveness of learning in nursing education; (6) research conducted in 2020 by Sugianto [10] showed the utilization of the discrepancy model used to evaluate individual learning programs at junior high school level. The findings obtained in Sugianto's research were two aspects of individual learning programs that were unsuitable to program standards. Those aspects include: (a) aspects of the preparation and organization, and (b) aspects of the implementation and assessment. Besides, the limitations found in Sugianto's research was that it had not shown a dominant indicator that was the main cause of the success of the program implementation. Based on the problems that occur in the field, the innovations that were initiated, as well as the results and limitations of some previous studies, it was necessary to conduct more in-depth research related to the development of a Discrepancy model based on Tat Twam Asi combined with TOPSIS calculations to get a dominant indicator triggering the effectiveness of the blended learning implementation (case studies at several IT vocational schools in Bali Province).

II. Basic Theory

A. Discrepancy Evaluation Model

The discrepancy is an evaluation model that is used to determine the comparison between actual performance that occurs and standards that have been previously set in the evaluation [11]. The discrepancy is an evaluation model that consists of four evaluation components, including definition, installation, process, and product [12]. Based on those several definitions of discrepancy, a general conclusion is that discrepancy is one of the evaluation models comparing work results with existing standards to obtain the level of discrepancy using four stages/components of evaluation, including definition, installation, process, and product.

B. Tat Twam Asi

According to Evitasari and Wiranti, "Tat Twam Asi" is one of the Balinese local wisdom that teaches equality in the behavior of every human being in establishing a relationship to create harmony [13]. According to Perbowosari, the term "Tat Twam Asi" means I am you. This contains the concept of togetherness. Four elements need to be built to maintain togetherness, including 1) having the same vision, 2) not being selfish, 3) being willing to sacrifice, and 4) being humble [14,15]. Based on some of those statements, the general conclusion is that Tat Twam Asi is a concept that was born from the philosophy of local Balinese wisdom which shows equality, and alignment in authority so that later it can lead to harmony and effectiveness in living life.

C. TOPSIS

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is one of the multi-criteria of the decision-making methods, the principle of which works to find alternative choices by taking into account the closest distance from the positive ideal solution and the farthest from the negative ideal solution to determine the relative closeness between the optimal solutions with an alternative [16]. The steps to search for alternative options using TOPSIS can be described as follows [17]:

1) Make a normalized decision matrix.

2) Make a normalized weighted decision matrix.

3) Determine the matrix for the positive ideal solution and the matrix for the negative ideal solution.

4) Determine the distance between the values of each alternative and the matrix for positive ideal solutions and the matrix for negative ideal solutions.

5) Determine the preference value for each alternative.

TOPSIS requires a performance rating of each Ai alternative on each normalized Cj criteria, by the following formula [18].

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \quad (1)$$

The positive ideal solution of A^+ and the negative ideal solution of A^- can be determined based on the normalized weight rating (y_{ij}), by the following formula [19].

$$y_{ij} = w_i r_{ij} \quad (2)$$

$$A^+ = \left(y_1^+, y_2^+, \cdots, y_n^+\right) \quad (3)$$

$$A^- = \left(y_1^-, y_2^-, \cdots, y_n^-\right) \quad (4)$$

Which:

$$y_j^+ = \begin{cases} \max_i y_{ij}; & \text{If j is the profit attribute} \\ \min_i y_{ij}; & \text{If j is a cost attribute} \end{cases}$$

$$y_j^- = \begin{cases} \min_i y_{ij}; & \text{If j is the profit attribute} \\ \max_i y_{ij}; & \text{If j is a cost attribute} \end{cases}$$

The distance between the A_i alternatives and the positive ideal solution is formulated as follows [20].

$$D_i^+ = \sqrt{\sum_{j=1}^{n} \left(y_i^+ - y_{ij}\right)^2} \quad (5)$$

The distance between the alternative A_i with a negative ideal solution is formulated as follows [21].

$$D_i^- = \sqrt{\sum_{j=1}^{n}\left(y_{ij} - y_i^-\right)^2} \tag{6}$$

The preference value for each alternative (V_i) is formulated as follows [22].

$$V_i = \frac{D_i^-}{D_i^- + D_i^+} \tag{7}$$

A greater value of V_i indicates that alternative A_i is preferred.

D. Blended Learning

Blended learning is a learning model that combines conventional learning that is carried out in the classroom and learning based on internet technology or other digital media, so that the learning process can be done quickly, easily, flexibly, and interaction between teachers and students through discussion in class and online outside the classroom [23]. Blended learning presents flexibility in place, time, and media used in the learning process without ignoring the elements of interaction that occur between teachers and students because the learning process can be done in the classroom or outside the classroom assisted by information technology [24,25]. Based on those statements, blended learning is a learning model that combines face-to-face learning directly in the room and outdoor learning assisted by information technology.

E. Discrepancy Evaluation Model based on Tat Twam Asi using TOPSIS Calculation

This model is a new breakthrough in developing the Discrepancy evaluation model that combines the concept of Tat Twam Asi with the TOPSIS method, making it easier to determine the dominant indicators that trigger the effectiveness of blended learning. The four components of discrepancy evaluation are given equal weight from evaluators based on the Tat Twam Asi concept reference then the weighting results are used in the TOPSIS calculation to obtain the preference value of each evaluation indicator so that later indicators can be obtained dominantly triggers the effectiveness of blended learning accurately.

III. Method

A. Research Approach

This research used a development approach by the Research and Development method. The research development model was Borg and Gall which consists of 10 stages of development [26-28], including: (1) research & field data collection; (2) planning; (3) design development; (4) initial trials; (5) revisions to the results of the initial trial; (6) field trial; (7) revision of the results of field trial; (8) usage trial; (9) final product revisions; (10) dissemination and implementation of the final product. Specifically, this paper focused on several stages undertaken to create a Tat Twam Asi-based Discrepancy evaluation model with TOPSIS calculations, including: (1) research & field data collection; (2) planning; (3) design development; (4) initial trial; and (5) revisions to initial trial results.

B. Research Subjects

The subjects involved in this research were two educational experts, two informatics experts, and 30 teachers, who would later be involved in conducting the initial trial. The education experts involved have a specific scientific field that was educational evaluation, while the informatics experts involved have a specific scientific field namely IT education.

C. Research Object

The object of research is the main topic that must be studied and solved. The object of this research was the design of a Discrepancy evaluation model based on Tat Twam Asi with TOPSIS calculation.

D. Research Location

The implementation of this research was located at IT vocational schools spread across six regencies in Bali. The six regencies include: Gianyar, Buleleng, Tabanan, Badung, Klungkung, and Denpasar.

E. Data Collection Instruments

Instruments used for collecting data in this research were in the form of questionnaires, photo documentation, and interview guidelines. The questionnaires were used to obtain primary data in the form of quantitative data from respondents as a basis for making decisions about the effectiveness percentage of the blended learning implementation. Interview guidelines were used to obtain secondary data as a basis for strengthening arguments qualitatively in supporting research findings. Photo documentation was used as proof that this research was indeed carried out and also used as valid evidence that showed the source of primary and secondary data obtained in this research.

F. Data Analysis Techniques

The technique used to analyze the data that had been collected was a quantitative descriptive technique through percentage descriptive calculation. The percentage descriptive calculation results were used as a basis for interpreting the research results on the Tat Twam Asi-based Discrepancy evaluation model. The percentage descriptive calculation is formulated as follows [29-35].

$$\text{Percentage } = \frac{\sum \text{(Answer} \times \text{Weight of Each Choice)}}{\text{n} \times \text{Highest Weight}} \times 100\% \tag{8}$$

Notes:

$\sum$ = Total; n = Number of all questionnaire items.

The percentage results were obtained from that formula and then converted into the eleven’s scale categorization. That categorization can be seen in Table I [36,37].

TABLE I. ELEVEN'S SCALE CATEGORIZATION

Effectiveness Percentage (%)	Category	Follow-up
95 to 100	Excellent	No needs revision
85 to 94	Very good	No needs revision
75 to 84	Good	No needs revision
65 to 74	More than enough	No needs revision
55 to 64	Enough	Revision
45 to 54	Almost enough	Revision
35 to 44	Minus	Revision
25 to 34	Very minus	Revision
15 to 24	Poor	Revision
5 to 14	Very poor	Revision
0 to 4	Highly poor	Revision

IV. RESULTS AND DISCUSSION

At the research stage and field data collection, several results were obtained, including aspects of evaluation standards, evaluation results in the field, and the weight of decision-makers. The full aspects related to standard evaluation can be seen in Table II, the evaluation results in the field can be seen in Table III, and the weight of decision-makers can be seen in Table IV.

Based on the data shown in Table III, there appears to be an imbalance that occurs between the percentage of effectiveness in field evaluation with the established effectiveness standards. Positive inequality occurs if the percentage of effectiveness in the field was higher than the percentage of effectiveness standards. Otherwise, if the percentage of effectiveness in the field was lower than percentage of effectiveness standards then negative inequality occurs. From Table III, several indicators were classified as negative inequality, including: indicators 11, 12, 13, 16, 18, 21, 22, 23, and 27. Indicators classified as positive inequality, including: indicators 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 14, 15, 17, 19, 20, 24, 25, 26, 28, 29, 30, and 31.

TABLE II. THE EVALUATION STANDARD ASPECTS OF THE BLENDED LEARNING IMPLEMENTATION IN SEVERAL VOCATIONAL SCHOOLS IN BALI PROVINCE WHICH REFERRED TO THE DISCREPANCY MODEL

Evaluation Components	Aspects/Criteria of Evaluation		Indicators		Percentage of Effectiveness Standards
Definition	C1	The legality of conducting blended learning	I-1	Education service regulation regarding the needs of blended learning	88
			I-2	Principals' regulation regarding the implementation of blended learning	90
	C2	Academics support	I-3	Principals' agreement	88
			I-4	Developer team support	90
			I-5	Teacher enthusiasm	85
			I-6	Students enthusiasm	85
	C3	Community support	I-7	Support from board of trustees / school committees	87
			I-8	Support from students' parents	87
Installation	C4	Management team readiness	I-9	Suitability of academic qualifications and scientific fields of the management team	88
			I-10	Management team competence	88
	C5	Facility and infrastructure readiness	I-11	Availability of hardware with adequate specifications	88
			I-12	Availability of software / platforms that suit on the needs	88
			I-13	Adequate internet access availability	90
			I-14	Availability of supporting physical infrastructure (such as tables, chairs, air conditioners, LCD projectors, etc.) that are still suitable for use	87
	C6	User competency readiness	I-15	The ability of teachers in operating computers and accessing the internet	86

			I-16	The ability of teachers to prepare digital teaching materials to support blended learning	86
			I-17	Students' expertise in operating computers and accessing the internet	86
Processes	C7	The socialization of the procedures for using blended learning	I-18	There was a socialization to teachers about the procedures for making digital teaching materials	87
			I-19	There was a socialization of the use of blended learning to teachers and students	87
	C8	Implementation of learning using blended learning	I-20	The implementation time of learning is in accordance with the time agreed upon by students and teachers	88
			I-21	The quality of material transferred by the teacher through blended learning can be easily understood by students	88
Product	The effectiveness of implementing blended learning from several dimensions:				
	C9	Tangibles	I-22	The condition of the classroom/ lab that is used in the organization of blended learning	88
			I-23	The condition of digital teaching materials that is used in the learning process based on blended learning	88
	C10	Reliability	I-24	Speed in accessing a blended learning platform	88
			I-25	Ease of operating a blended learning platform	88
	C11	Responsiveness	I-26	Platform speed in responding to the process of data manipulation (input, edit, and delete digital teaching materials) into blended learning	87
			I-27	The speed of response given by the teacher when discussing with students through blended learning	86
	C12	Assurance	I-28	Security guarantees questions/tests which are provided by teachers in blended learning	90
			I-29	The security guarantee of each task deposited by students into blended learning	90
	C13	Empathy	I-30	The availability of facilities for giving advice/ complaints from students to the learning process through blended learning	90
			I-31	The availability of feedback facilities from teachers on existing suggestions/ complaints related to the learning process through blended learning	90

TABLE III. FIELD EVALUATION RESULTS REFERRING TO THE DISCREPANCY MODEL OF BLENDED LEARNING IMPLEMENTATION IN SEVERAL IT VOCATIONAL SCHOOLS IN BALI PROVINCE

Code of Indicators	**Percentage of Effectiveness Standards**	**Percentage of Effectiveness in Field Evaluation (%)**	***Discrepancy***
I-1	88.000	91.765	3.765
I-2	90.000	92.353	2.353
I-3	88.000	91.176	3.176
I-4	90.000	90.588	0.588
I-5	85.000	85.294	0.294
I-6	85.000	85.882	0.882
I-7	87.000	88.824	1.824
I-8	87.000	89.412	2.412
I-9	88.000	88.235	0.235
I-10	88.000	88.824	0.824
I-11	88.000	80.588	-7.412
I-12	88.000	85.294	-2.706
I-13	90.000	85.882	-4.118
I-14	87.000	87.647	0.647
I-15	86.000	86.471	0.471
I-16	86.000	80.588	-5.412
I-17	86.000	87.059	1.059
I-18	87.000	75.294	-11.706
I-19	87.000	87.059	0.059
I-20	88.000	88.235	0.235
I-21	88.000	86.176	-6.824
I-22	88.000	86.471	-1.529
I-23	88.000	82.353	-5.647
I-24	88.000	88.235	0.235
I-25	88.000	88.824	0.824
I-26	87.000	88.235	1.235
I-27	86.000	84.118	-1.882
I-28	90.000	91.765	1.765
I-29	90.000	91.176	1.176
I-30	90.000	92.353	2.353
I-31	90.000	92.941	2.941
Average		**87.230**	

The data in Table IV shows the weighted value given by experts for each evaluation criteria. The weight value given to each evaluation criteria refers to the Tat Twam Asi concept. Tat Twam Asi was a concept that prioritizes equality/similarity of authority for each expert in providing a weighting assessment of each evaluation criteria. Therefore an average weight score calculation was performed to achieve the same authority of each expert. The weighted average results were then divided by the total number of weighted average, so we got a weight value that refers to Tat Twam Asi for each evaluation criteria. There were 14 evaluation criteria that were given weight referring to Tat Twam Asi.

TABLE IV. THE WEIGHTS GIVEN BY THE DECISION-MAKERS TO EACH EVALUATION CRITERIA REFERS TO THE TAT TWAM ASI CONCEPT

Code of Criteria	**Weights Given by Experts**				**Average of Weights**	**Weights Refers to Tat Twam Asi**
	Expert 1	**Expert 2**	**Expert 3**	**Expert 4**		
C1	5	4	5	5	4.75	0.080
C2	4	4	5	4	4.25	0.071
C3	4	5	4	4	4.25	0.071
C4	5	5	5	4	4.75	0.080
C5	4	4	4	5	4.25	0.071
C6	4	5	4	4	4.25	0.071
C7	4	4	5	4	4.25	0.071
C8	4	4	5	5	4.50	0.076
C9	4	3	4	4	3.75	0.063
C10	4	5	4	4	4.25	0.071
C11	4	5	5	4	4.50	0.076
C12	4	5	4	5	4.50	0.076
C13	4	5	5	5	4.75	0.080
C14	2	3	3	2	2.50	0.042
∑					59.50	1.000

Notes:
• C1 to C13 were evaluation criteria as mentioned earlier in Table III.
• C14 was specifically an *Discrepancy* criteria.

At the planning stage, activities and time were regulated, as well as personnel involved in developing the evaluation model design. The details of activities and time needed in developing the design of a Tat Twam Asi-based Discrepancy evaluation model with a TOPSIS calculation can be seen in Table V. Personnel arrangements can be seen in Table VI.

At the design development stage, was made conceptual design and user interface design of the evaluation model. The design of the Discrepancy evaluation model based on Tat Twan Asi with TOPSIS calculation in finding dominant indicators that determine the success of blended learning implementation can be seen in Fig. 1 and user interface design in Fig. 2.

TABLE V. ACTIVITIES DETAILS IN THE DEVELOPMENT OF TAT TWAM ASI-BASED ON DISCREPANCY EVALUATION MODEL DESIGN USING TOPSIS CALCULATION

No.	**Activities**	**Time (Day)**
1	Determination of evaluation components	2
2	Determination of evaluation criteria/aspects	2
3	Determination of evaluation indicators	2
3	Determination of weights for each criteria	2
4	Making initial design	3
5	Trial calculation of TOPSIS	2
6	Revised trial results	2
7	Making final design	2
	Total	17

TABLE VI. PERSONNEL DETAILS WERE INVOLVED IN THE DEVELOPMENT OF TAT TWAM ASI-BASED ON DISCREPANCY EVALUATION MODEL DESIGN WITH TOPSIS CALCULATIONS

No	Activities	Number of personnel	Information
1	Determination of evaluation components	3	1 Research chair and 2 Research members
2	Determination of evaluation criteria/aspects	3	1 Research chair and 2 Research members
3	Determination of evaluation indicators	3	1 Research chair and 2 Research members
3	Determination of weights for each criteria	4	2 Education experts and 2 Informatics experts
4	Making initial design	3	1 Research chair and 2 Research members
5	Trial calculation of TOPSIS	34	2 Education experts, 2 Informatics expert, and 30 teachers
6	Revised trial results	3	1 Research chair and 2 Research members
7	Making final design	3	1 Research chair and 2 Research members

Fig. 1 shows the Discrepancy evaluation model consists of four evaluation components, including definition, installation, process, and product. From each of the evaluation components, there were several indicators used to measure the effectiveness of the blended learning implementation at the IT vocational schools. In Fig. 1, the evaluation indicators were displayed in the form of an indicator code. For more details about the description of each indicator's code can be seen in Table II. The results of the effectiveness percentage obtained in the field were then compared with the percentage of effectiveness standards. The inequality that occurs due to the process of comparing the effectiveness percentage, then it was used as one of the evaluation criteria from a total of 14 existing evaluation criteria. Those fourteen criteria were then given a weight value obtained from the weighting process of experts referring to the Tat Twam Asi concept. Based on the weight value given by the experts to each evaluation criteria, then the calculation process can be performed using the TOPSIS formula to determine the most dominant indicators as triggers for the success or effectiveness of blended learning implementation.

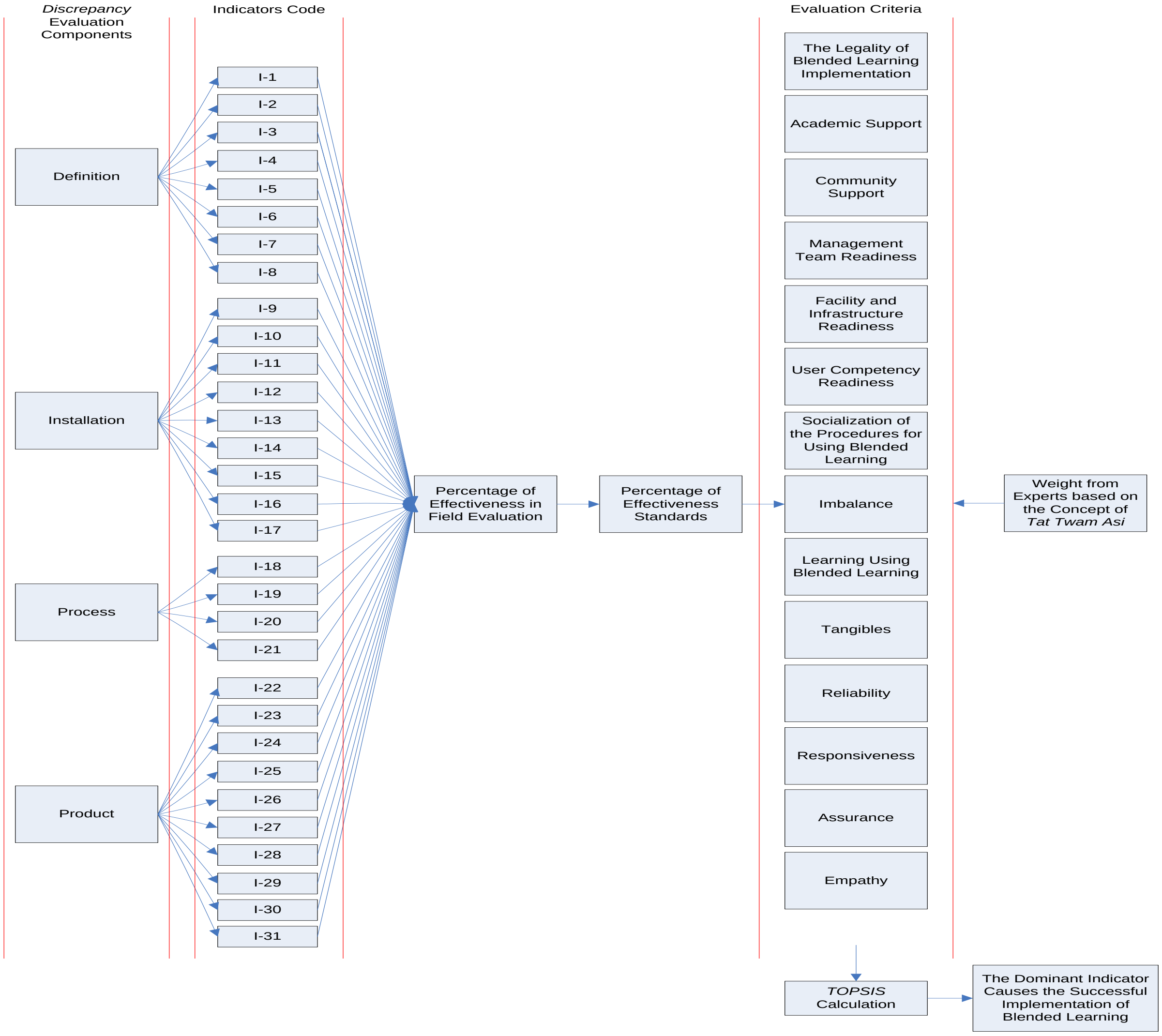
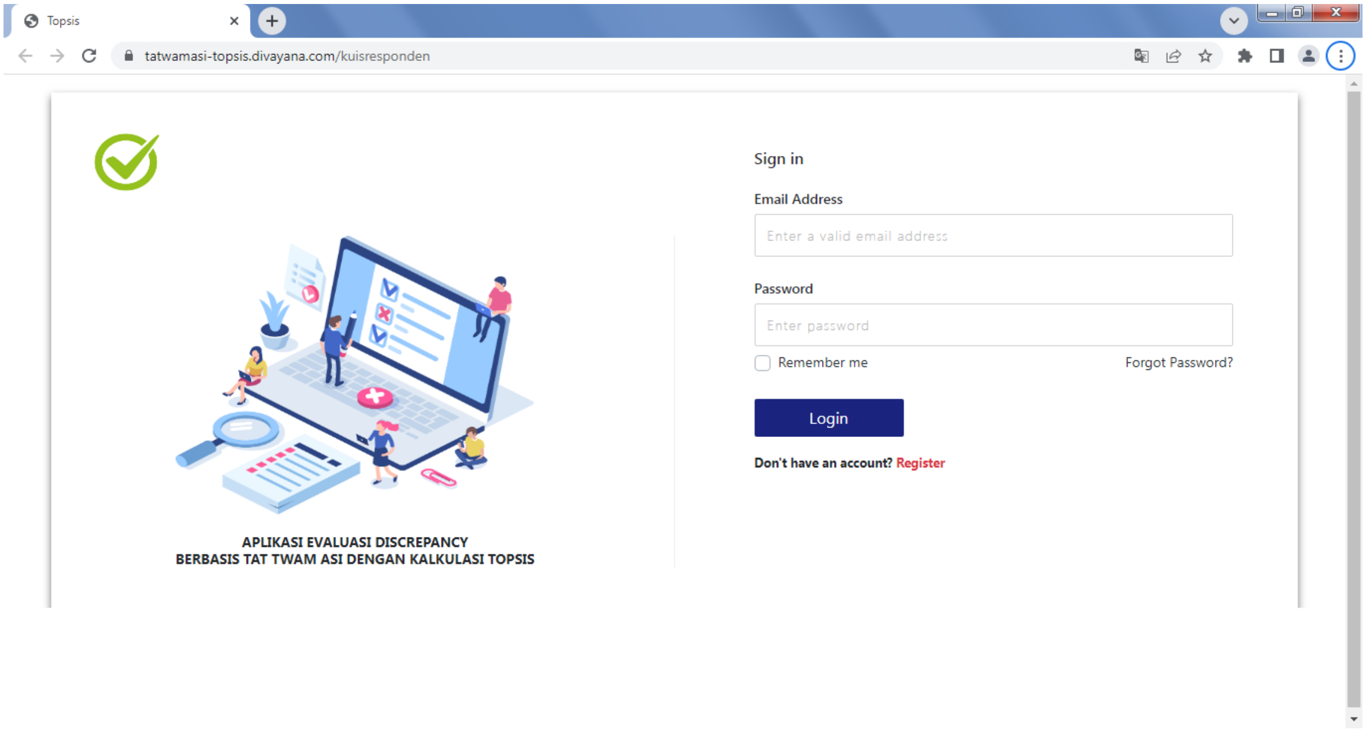
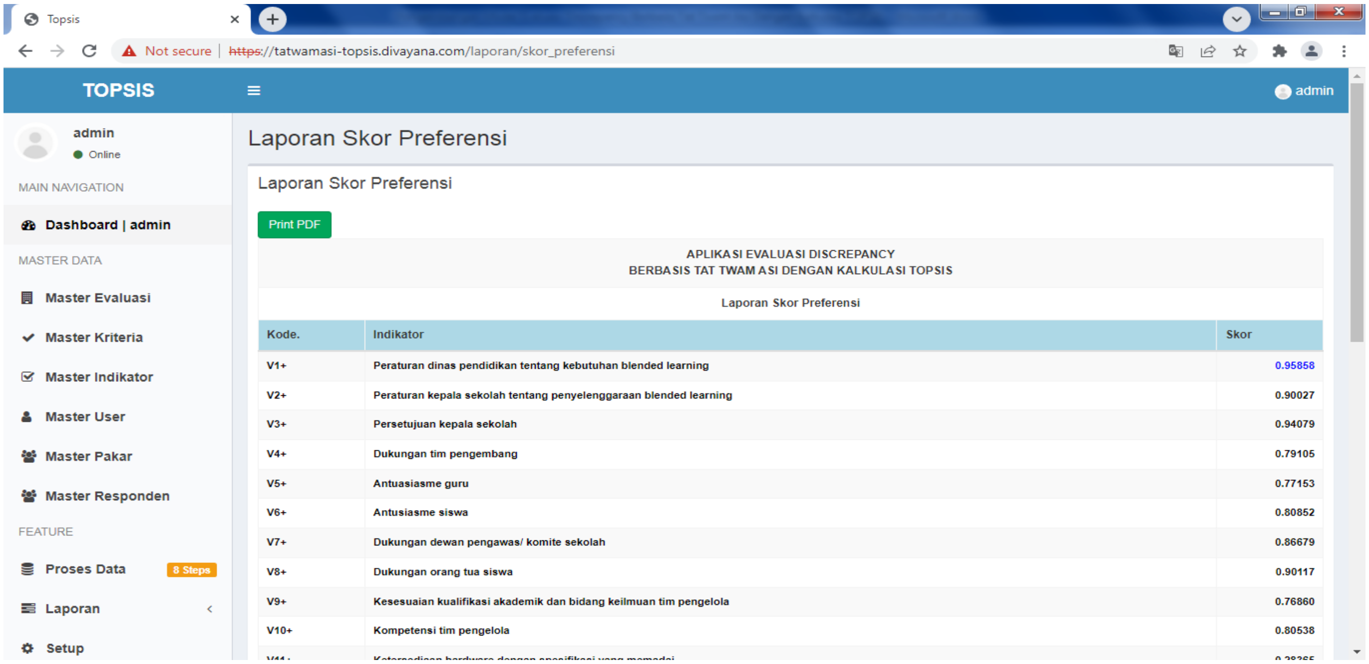

Discrepancy Evaluation Components
Indicators Code
Evaluation Criteria
Definition
Installation
Process
Product
I-1
I-2
I-3
I-4
I-5
I-6
I-7
I-8
I-9
I-10
I-11
I-12
I-13
I-14
I-15
I-16
I-17
I-18
I-19
I-20
I-21
I-22
I-23
I-24
I-25
I-26
I-27
I-28
I-29
I-30
I-31
Percentage of Effectiveness in Field Evaluation
Percentage of Effectiveness Standards
The Legality of Blended Learning Implementation
Academic Support
Community Support
Management Team Readiness
Facility and Infrastructure Readiness
User Competency Readiness
Socialization of the Procedures for Using Blended Learning
Imbalance
Learning Using Blended Learning
Tangibles
Reliability
Responsiveness
Assurance
Empathy
Weight from Experts based on the Concept of Tat Twam Asi
TOPSIS Calculation
The Dominant Indicator Causes the Successful Implementation of Blended Learning

Fig. 1. Conceptual Design of Discrepancy Evaluation Model based on Tat Twam Asi with TOPSIS Calculation in Finding Dominant Indicators that Determine the Success of Blended Learning Implementation.


Topsis
tatwamasi-topsis.divayana.com/kuisresponden
Sign in
Email Address
Enter a valid email address
Password
Enter password
□ Remember me
Forgot Password?
Login
Don't have an account? Register

APLIKASI EVALUASI DISCREPANCY
BERBASIS TAT TWAM ASI DENGAN KALKULASI TOPSIS



Laporan Skor Preferensi

Laporan Skor Preferensi

APLIKASI EVALUASI DISCREPANCY
BERBASIS TAT TWAM ASI DENGAN KALKULASI TOPSIS

Laporan Skor Preferensi

Kode.	Indikator	Skor
V1+	Peraturan dinas pendidikan tentang kebutuhan blended learning	0.95858
V2+	Peraturan kepala sekolah tentang penyelenggaraan blended learning	0.90027
V3+	Persetujuan kepala sekolah	0.94079
V4+	Dukungan tim pengembang	0.79105
V5+	Antuasiasme guru	0.77153
V6+	Antusiasme siswa	0.80852
V7+	Dukungan dewan pengawas/ komite sekolah	0.86679
V8+	Dukungan orang tua siswa	0.90117
V9+	Kesesuaian kualifikasi akademik dan bidang keilmuan tim pengelola	0.76860
V10+	Kompetensi tim pengelola	0.80538
V11+	Ketersediaan hardware dengan spesifikasi yang memadai	0.38365

Fig. 2. User Interface Design of Discrepancy Evaluation Model based on Tat Twam Asi with TOPSIS Calculation in Finding Dominant Indicators that determine the Success of Blended Learning Implementation (Indonesian Version).

In the initial testing phase toward the accuracy of the use of the TOPSIS method in the evaluation model design, a simulation calculation of the TOPSIS formula was performed to determine the dominant indicator. There were two education experts, two informatics experts, and 30 teachers involved in the TOPSIS calculation simulation process. The data which was used in the complete TOPSIS calculation simulation can be seen in Table VII.

From the data shown in Table VII, it can be explained that the scores given by the black block in columns C1 to C13 were obtained from the percentage value of effectiveness in field evaluation for each evaluation indicator (as shown in Table III). Unblocked scores were obtained from the average value of the percentage of effectiveness in field evaluation (as shown in Table III). The score entered in column C14 (indicated by a gray block) was obtained from the inequality score also shown in Table III.

TABLE VII. PRELIMINARY DATA FOR SIMULATION OF TOPSIS CALCULATION

Code of Indicators	**Criteria**													
	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**
I-1	91.765	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	3.765
I-2	92.353	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	2.353
I-3	87.230	91.176	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	3.176
I-4	87.230	90.588	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	0.588
I-5	87.230	85.294	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	0.294
I-6	87.230	85.882	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	0.882
I-7	87.230	87.230	88.824	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	1.824
I-8	87.230	87.230	89.412	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	2.412
I-9	87.230	87.230	87.230	88.235	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	0.235
I-10	87.230	87.230	87.230	88.824	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	0.824
I-11	87.230	87.230	87.230	87.230	80.588	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	-7.412
I-12	87.230	87.230	87.230	87.230	85.294	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	-2.706
I-13	87.230	87.230	87.230	87.230	85.882	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	-4.118
I-14	87.230	87.230	87.230	87.230	87.647	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	0.647
I-15	87.230	87.230	87.230	87.230	87.230	86.471	87.230	87.230	87.230	87.230	87.230	87.230	87.230	0.471
I-16	87.230	87.230	87.230	87.230	87.230	80.588	87.230	87.230	87.230	87.230	87.230	87.230	87.230	-5.412
I-17	87.230	87.230	87.230	87.230	87.230	87.059	87.230	87.230	87.230	87.230	87.230	87.230	87.230	1.059
I-18	87.230	87.230	87.230	87.230	87.230	87.230	75.294	87.230	87.230	87.230	87.230	87.230	87.230	-11.706
I-19	87.230	87.230	87.230	87.230	87.230	87.230	87.059	87.230	87.230	87.230	87.230	87.230	87.230	0.059
I-20	87.230	87.230	87.230	87.230	87.230	87.230	87.230	88.235	87.230	87.230	87.230	87.230	87.230	0.235
I-21	87.230	87.230	87.230	87.230	87.230	87.230	87.230	86.176	87.230	87.230	87.230	87.230	87.230	-6.824
I-22	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	86.471	87.230	87.230	87.230	87.230	-1.529
I-23	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	82.353	87.230	87.230	87.230	87.230	-5.647
I-24	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	88.235	87.230	87.230	87.230	0.235
I-25	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	88.824	87.230	87.230	87.230	0.824
I-26	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	88.235	87.230	87.230	1.235
I-27	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	84.118	87.230	87.230	-1.882
I-28	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	91.765	87.230	1.765
I-29	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	91.176	87.230	1.176
I-30	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	92.353	2.353
I-31	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	87.230	92.941	2.941

Based on the data shown in Table VII, the TOPSIS calculation process was then performed by following the steps as follows:

1) Determine the normalized matrix using the formula shown earlier in equation (1)

$$|x_1| = \sqrt{91.765^2 + 92.353^2 + 29(87.230)^2} = 487.454$$

$$r_{11} = \frac{x_{11}}{|x_1|} = \frac{91.765}{487.454} = 0.1883$$

$$r_{21} = \frac{x_{21}}{|x_1|} = \frac{92.353}{487.454} = 0.1895$$

$$r_{31} = \frac{x_{31}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{41} = \frac{x_{41}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{51} = \frac{x_{51}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{61} = \frac{x_{61}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{71} = \frac{x_{71}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{81} = \frac{x_{81}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{91} = \frac{x_{91}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{101} = \frac{x_{101}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{111} = \frac{x_{111}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{121} = \frac{x_{121}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{131} = \frac{x_{131}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{141} = \frac{x_{141}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{151} = \frac{x_{151}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{161} = \frac{x_{161}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{171} = \frac{x_{171}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{181} = \frac{x_{181}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{191} = \frac{x_{191}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{201} = \frac{x_{201}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{211} = \frac{x_{211}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{221} = \frac{x_{221}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{231} = \frac{x_{231}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{241} = \frac{x_{241}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{251} = \frac{x_{251}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{261} = \frac{x_{261}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{271} = \frac{x_{271}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{281} = \frac{x_{281}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{291} = \frac{x_{291}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{301} = \frac{x_{301}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

$$r_{311} = \frac{x_{311}}{|x_1|} = \frac{87.230}{487.454} = 0.1789$$

same calculation up to $|x_{14}|$

$$|x_{14}| = \sqrt{3.765^2 + 2.353^2 + 3.176^2 + 0.588^2 + 0.294^2 + 0.882^2 + 1.824^2 +}$$

$$\sqrt{+2.412^2 + 0.235^2 + 0.824^2 + (-7.412)^2 + (-2.706)^2 + (-4.118)^2 +}$$

$$\sqrt{+0.647^2 + 0.471^2 + (-5.412)^2 + 1.059^2 + (-11.706)^2 + 0.059^2 +}$$

$$\sqrt{+0.235^2 + (-6.824)^2 + (-1.529)^2 + (-5.647)^2 + 0.235^2 + 0.824^2 +}$$

$$\sqrt{+1.235^2 + (-1.882)^2 + 1.765^2 + +1.176^2 + 2.353^2 + (2.941)^2} = 19.837$$

$$r_{114} = \frac{x_{114}}{|x_{14}|} = \frac{3.765}{19.837} = 0.1898$$

$$r_{214} = \frac{x_{214}}{|x_{14}|} = \frac{2.353}{19.837} = 0.1186$$

$$r_{314} = \frac{x_{314}}{|x_{14}|} = \frac{3.176}{19.837} = 0.1601$$

$$r_{414} = \frac{x_{414}}{|x_{14}|} = \frac{0.588}{19.837} = 0.0296$$

$$r_{514} = \frac{x_{514}}{|x_{14}|} = \frac{0.294}{19.837} = 0.0148$$

$$r_{614} = \frac{x_{614}}{|x_{14}|} = \frac{0.882}{19.837} = 0.0445$$

$$r_{714} = \frac{x_{714}}{|x_{14}|} = \frac{1.824}{19.837} = 0.0919$$

$$r_{814} = \frac{x_{814}}{|x_{14}|} = \frac{2.412}{19.837} = 0.1216$$

$$r_{914} = \frac{x_{914}}{|x_{14}|} = \frac{0.235}{19.837} = 0.0118$$

$$r_{1014} = \frac{x_{1014}}{|x_{14}|} = \frac{0.824}{19.837} = 0.0415$$

$$r_{1114} = \frac{x_{1114}}{|x_{14}|} = \frac{-7.412}{19.837} = -0.3736$$

$$r_{1214} = \frac{x_{1214}}{|x_{14}|} = \frac{-2.706}{19.837} = -0.1364$$

$$r_{1314} = \frac{x_{1314}}{|x_{14}|} = \frac{-4.118}{19.837} = -0.2076$$

$$r_{1414} = \frac{x_{1414}}{|x_{14}|} = \frac{0.647}{19.837} = 0.0326$$

$$r_{1514} = \frac{x_{1514}}{|x_{14}|} = \frac{0.471}{19.837} = 0.0237$$

$$r_{1614} = \frac{x_{1614}}{|x_{14}|} = \frac{-5.412}{19.837} = -0.2728$$

$$r_{1714} = \frac{x_{1714}}{|x_{14}|} = \frac{1.059}{19.837} = 0.0534$$

$$r_{1814} = \frac{x_{1814}}{|x_{14}|} = \frac{-11.706}{19.837} = -0.5901$$

$$r_{1914} = \frac{x_{1914}}{|x_{14}|} = \frac{0.059}{19.837} = 0.0030$$

$$r_{2014} = \frac{x_{2014}}{|x_{14}|} = \frac{0.235}{19.837} = 0.0118$$

$$r_{2114} = \frac{x_{2114}}{|x_{14}|} = \frac{-6.824}{19.837} = -0.3440$$

$$r_{2214} = \frac{x_{2214}}{|x_{14}|} = \frac{-1.529}{19.837} = -0.0771$$

$$r_{2314} = \frac{x_{2314}}{|x_{14}|} = \frac{-5.647}{19.837} = -0.2847$$

$$r_{2414} = \frac{x_{2414}}{|x_{14}|} = \frac{0.235}{19.837} = 0.0118$$

$$r_{2514} = \frac{x_{2514}}{|x_{14}|} = \frac{0.824}{19.837} = 0.0415$$

$$r_{2614} = \frac{x_{2614}}{|x_{14}|} = \frac{1.235}{19.837} = 0.0623$$

$$r_{2714} = \frac{x_{2714}}{|x_{14}|} = \frac{-1.882}{19.837} = -0.0949$$

$$r_{2814} = \frac{x_{2814}}{|x_{14}|} = \frac{1.765}{19.837} = 0.0890$$

$$r_{2914} = \frac{x_{2914}}{|x_{14}|} = \frac{1.176}{19.837} = 0.0593$$

$$r_{3014} = \frac{x_{3014}}{|x_{14}|} = \frac{2.353}{19.837} = 0.1186$$

$$r_{3114} = \frac{x_{3114}}{|x_{14}|} = \frac{2.941}{19.837} = 0.1483$$

2) Determine the R matrix: The results of the normalization were plotted into the 31 × 14. R matrix with the intended normalized matrix can be seen in Figure 3.

$$R=\begin{bmatrix}
0.1883 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.1898 \\
0.1895 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.1186 \\
0.1789 & 0.1874 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.1601 \\
0.1789 & 0.1862 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0296 \\
0.1789 & 0.1753 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0148 \\
0.1789 & 0.1766 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0445 \\
0.1789 & 0.1793 & 0.1826 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0919 \\
0.1789 & 0.1793 & 0.1838 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.1216 \\
0.1789 & 0.1793 & 0.1794 & 0.1815 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0118 \\
0.1789 & 0.1793 & 0.1794 & 0.1827 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0415 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1665 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & -0.3736 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1762 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & -0.1364 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1774 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & -0.2076 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1811 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0326 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1785 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0237 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1664 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & -0.2728 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1797 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0534 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1557 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & -0.5901 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1800 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0030 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1817 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & 0.0118 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1774 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & -0.3440 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1784 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & -0.0771 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1699 & 0.1794 & 0.1797 & 0.1790 & 0.1789 & -0.2847 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1815 & 0.1797 & 0.1790 & 0.1789 & 0.0118 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1827 & 0.1797 & 0.1790 & 0.1789 & 0.0415 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1818 & 0.1790 & 0.1789 & 0.0623 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1733 & 0.1790 & 0.1789 & -0.0949 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1883 & 0.1789 & 0.0890 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1871 & 0.1789 & 0.0593 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1894 & 0.1186 \\
0.1789 & 0.1793 & 0.1794 & 0.1794 & 0.1802 & 0.1801 & 0.1804 & 0.1796 & 0.1800 & 0.1794 & 0.1797 & 0.1790 & 0.1906 & 0.1483
\end{bmatrix}$$

Fig. 3. R Matrix.

3) Determine the Y matrix: After the R matrix was obtained, then calculation was performed to determine the Y matrix. The Y matrix was the weighted normalized matrix. The way to get the Y matrix was to do a multiplication of the R matrix with expert weights referring to the Tat Twam Asi concept shown in Table IV. In general, the matrix multiplication to determine the Y matrix can be written as follows.

$$Y = [R] \times [0.080\ \ 0.071\ \ 0.071\ \ 0.080\ \ 0.071\ \ 0.071\ \ 0.076\ \ 0.063\ \ 0.076\ \ 0.076\ \ 0.076\ \ 0.080\ \ 0.042]$$

The complete results of the calculation of determining the Y matrix can be seen in Fig. 4.

$$Y = \begin{bmatrix}
0.01506 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00797 \\
0.01516 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00498 \\
0.01431 & 0.01331 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00672 \\
0.01431 & 0.01322 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00124 \\
0.01431 & 0.01245 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00062 \\
0.01431 & 0.01254 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00187 \\
0.01431 & 0.01273 & 0.01296 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00386 \\
0.01431 & 0.01273 & 0.01305 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00511 \\
0.01431 & 0.01273 & 0.01274 & 0.01452 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00050 \\
0.01431 & 0.01273 & 0.01274 & 0.01462 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00174 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01182 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & -0.01569 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01251 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & -0.00573 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01260 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & -0.00872 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01286 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00137 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01267 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00100 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01181 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & -0.01146 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01276 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00224 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01105 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & -0.02478 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01278 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00013 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01381 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & 0.00050 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01348 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & -0.01445 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01124 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & -0.00324 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01070 & 0.01274 & 0.01366 & 0.01360 & 0.01431 & -0.01196 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01289 & 0.01366 & 0.01360 & 0.01431 & 0.00050 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01297 & 0.01366 & 0.01360 & 0.01431 & 0.00174 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01382 & 0.01360 & 0.01431 & 0.00262 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01317 & 0.01360 & 0.01431 & -0.00399 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01431 & 0.01431 & 0.00374 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01422 & 0.01431 & 0.00249 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01515 & 0.00498 \\
0.01431 & 0.01273 & 0.01274 & 0.01435 & 0.01279 & 0.01279 & 0.01281 & 0.01365 & 0.01134 & 0.01274 & 0.01366 & 0.01360 & 0.01525 & 0.00623
\end{bmatrix}$$

Fig. 4. Y Matrix.

4) Determine the matrix for positive ideal solutions and the matrix for negative ideal solutions.

The matrix for positive and negative ideal solutions was strongly influenced by the classification of each evaluation criteria. The fourteen evaluation criteria in this research were classified as profit attributes. Based on those, it can be calculated the matrix for positive ideal solutions and the matrix for negative ideal solutions as follows.

a) The Matrix for Positive Ideal Solutions

y_1^+= max{0.01506; 0.01516; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431} = 0.01506

↓ same calculation up to y_{14}^+

y_{14}^+=max{0.00797; 0.00498; 0.00672; 0.00124; 0.00062; 0.00187; 0.00386; 0.00511; 0.00050; 0.00174; -0.01569; -0.00573; -0.00872; 0.00137; 0.00100; -0.01146; 0.00224; -0.02478; 0.00013; 0.00050; -0.01445; -0.00324; -0.01196; 0.00050; 0.00174; 0.00262; -0.00399; 0.00374; 0.00249; 0.00498; 0.00623} = 0.00797

A^+ = {0.01516; 0.01331; 0.01305; 0.01462; 0.01286; 0.01279; 0.01281; 0.01381; 0.01134; 0.01297; 0.01382; 0.01431; 0.01525; 0.00797}

b) The Matrix for Negative Ideal Solutions

y_1^- = min{0.01506; 0.01516; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431; 0.01431} = 0.01431

y_2^- = min{0.01273; 0.01273; 0.01331; 0.01322; 0.01245; 0.01254; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273; 0.01273} = 0.01245

↓ same calculation up to y_{14}^-

y_{14}^- =min{0.00797; 0.00498; 0.00672; 0.00124; 0.00062; 0.00187; 0.00386; 0.00511; 0.00050; 0.00174; -0.01569; -0.00573; -0.00872; 0.00137; 0.00100; -0.01146; 0.00224; -0.02478; 0.00013; 0.00050; -0.01445; -0.00324; -0.01196; 0.00050; 0.00174; 0.00262; -0.00399; 0.00374; 0.00249; 0.00498; 0.00623} = -0.02478

A^- = {0.01431; 0.01245; 0.01274; 0.01435; 0.01182; 0.01181; 0.01105; 0.01348; 0.01070; 0.01274; 0.01317; 0.01360; 0.01431; -0.02478}

5) Determine the distance between the scores of each indicator with the matrix for a positive ideal solution and a negative ideal solution.

a) The distance between the scores of each indicator and the matrix for a positive ideal solution.

$$D_1^+ = \sqrt{(0.01506-0.01516)^2+(0.01273-0.01331)^2+(0.01274-0.01305)^2+(0.01435-0.01462)^2+} \ldots$$

$$\sqrt{..+(0.01279-0.01286)^2+(0.01279-0.01279)^2+(0.01281-0.01281)^2+(0.01365-0.01381)^2+\ \ldots}$$

$$\sqrt{..+(0.01134-0.01134)^2+(0.01274-0.01297)^2+(0.01366-0.01382)^2+(0.01360-0.01431)^2+} \ldots$$

$$\sqrt{..+(0.01431-0.01525)^2+(0.00797-0.00797)^2}$$

$= 0.00141$

↓ same calculation up to D_{31}^+

$$D_{31}^+ = \sqrt{(0.01431-0.01516)^2+(0.01273-0.01331)^2+(0.01274-0.01305)^2+(0.01435-0.01462)^2+} \ldots$$

$$\sqrt{..+(0.01279-0.01286)^2+(0.01279-0.01279)^2+(0.01281-0.01281)^2+(0.01365-0.01381)^2+\ \ldots}$$

$$\sqrt{..+(0.01134-0.01134)^2+(0.01274-0.01297)^2+(0.01366-0.01382)^2+(0.01360-0.01431)^2+} \ldots$$

$$\sqrt{..+(0.01525-0.01525)^2+(0.00623-0.00797)^2}$$

$= 0.00221$

b) The distance between the scores of each indicator and the matrix for negative ideal solutions

$$D_1^- = \sqrt{(0.01506-0.01431)^2+(0.01273-0.01245)^2+(0.01274-0.01274)^2+(0.01435-0.01435)^2+} \ldots$$

$$\sqrt{..+(0.01279-0.01182)^2+(0.01279-0.01181)^2+(0.01281-0.01105)^2+(0.01365-0.01348)^2+\ \ldots}$$

$$\sqrt{..+(0.01134-0.01070)^2+(0.01274-0.01274)^2+(0.01366-0.01317)^2+(0.01360-0.01360)^2+} \ldots$$

$$\sqrt{..+(0.01431-0.01431)^2+(0.00797-(-0.02478))^2}$$

= 0.03285

↓ same calculation up to D_{31}^-

$$D_{31}^- = \sqrt{(0.01431-0.01431)^2+(0.01273-0.01245)^2+(0.01274-0.01274)^2+(0.01435-0.01435)^2+\ldots}$$
$$\sqrt{..+(0.01279-0.01182)^2+(0.01279-0.01181)^2+(0.01281-0.01105)^2+(0.01365-0.01348)^2+\ \ldots}$$
$$\sqrt{..+(0.01134-0.01070)^2+(0.01274-0.01274)^2+(0.01366-0.01317)^2+(0.01360-0.01360)^2+\ldots}$$
$$\sqrt{..+(0.01525-0.01431)^2+(0.00623-(-0.02478))^2}$$
$$= 0.03112$$

6) Determine the preference score for each indicator.

$$V_1 = \frac{D_1^-}{D_1^- + D_1^+}$$
$$= \frac{0.03285}{0.03285+0.00141}$$
$$= 0.95880$$

↓ same calculation up to V_{31}

$$V_{31} = \frac{D_{31}^-}{D_{31}^- + D_{31}^+}$$
$$= \frac{0.03112}{0.03112+0.00221}$$
$$= 0.93381$$

7) Make decisions based on preference scores: Based on the results of the preference score, the most dominant indicator as a trigger for the success of blended learning implementation at IT vocational schools was V1, namely, I1 (the education service regulation regarding the need for blended learning).

At the revision stage of the initial trial results, there was nothing that needs to be revised in major related to the TOPSIS calculation process used in the design of the evaluation model. In addition, based on the results of the field evaluation shown in Table III, the effectiveness percentage was 87.230%. If that effectiveness percentage was matched with eleven's scale categorization (as shown in Table I) shows that the evaluation model design was included in the excellent category. So, In general, there was nothing that needs to revise in the evaluation model design. Therefore it can be concluded in general that the evaluation model design has been made optimally and the calculation simulation of the TOPSIS formula has been able to show an accurate calculation in determining the most dominant indicator as a trigger for the success of blended learning implementation at IT vocational schools.

Ma and Lee's research has similarities with this study in evaluating the effectiveness of blended learning implementation. However, the difference is in the uses of the evaluation model. Ma and Lee's research [38] used the ARCS (Attention, Relevance, Confidence, and Satisfaction) model, while this study in principle used the Discrepancy model. The limitation of Ma and Lee's research was it had not shown an accurate calculation process in determining the trigger indicators for the effectiveness of the blended learning implementation. Martín-Martínez et al.'s research have similarities with this study related to the object being evaluated, the difference is the evaluation mechanism. This study used a combination of the Discrepancy model, the TOPSIS method, and the Tat Twam Asi concept to determine the trigger indicators for the effectiveness of the blended learning implementation. The research of Martín-Martínez et al. [39] used a matrix of rotated factors to determine the level of effectiveness of the blended learning implementation. Sukirman et al.'s research have similarities with this study related to measuring the effectiveness of the blended learning implementation. The difference is the research approach. This study used an evaluative approach, while Sukirman et al.'s research [40] used an experimental study approach.

In general, this research has been able to give contributions to answering some of the limitations previously found in Embi et al.'s research, Istanbul and Supriadi's research, Agustina and Mukhtaruddin's research, Ngala et al.'s research, Siswadi et al.'s research, and Sugianto's research, through showing the dominant indicators that were the main cause of the successful implementation of the program (in this case blended learning at IT vocational schools in Bali Province) using the Discrepancy evaluation model based on Tat Twam Asi with TOPSIS calculation. Although this research was felt to solve the limitations of some previous studies, this research was also not perfect. There were several things found as limitations in this research, including 1) the indicators related to the readiness of funds in realizing blended learning have not been discussed in detail and 2) indicators related to the governance of the use of funds incurred in the blended learning administration have not been discussed in detail

V. Conclusion

The design of the Discrepancy evaluation model based on Tat Twam Asi with the TOPSIS calculation developed through this research was able to show the stages of structured evaluation and through an accurate calculation process in determining the dominant indicators that trigger the effectiveness/success of blended learning implementation in IT vocational schools in Bali Province. The evaluation stages used in this model were based on the Discrepancy model which has four evaluation components, including definition, installation, process, and product. The calculation process to determine the dominant indicators in this evaluation model uses the TOPSIS formula with the average weight given by experts for each criteria referring to the Tat Twam Asi concept that prioritizes equality of rights/authority. Future work that can be done to overcome the limitations found in this research is to include indicators of the readiness of funds in the installation component and include indicators of funding governance in the process components contained in the evaluation model.

Acknowledgment

The authors would like thank to the Directorate General of Research and Development, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia which provided funding for this research through research grant number: 1219/UN48.16/LT/2022. The authors also would like to thank for Rector and Chair of the Research and Community Service Institute of Universitas Pendidikan Ganesha who provided the opportunity, guidance, and support to complete this research on time.

References

[1] P. Mozelius, and C. Rydell, “Problems Affecting Successful Implementation of Blended Learning in Higher Education - The Teacher Perspective,” *International Journal of Information and Communication Technologies in Education*, Vol. 6, No.1, pp. 4–13, 2017.

[2] P. W. A. Suyasa, P. S. Kurniawan, I. P. W. Ariawan, W. Sugandini, N. D. M. S. Adnyawati, I. D. A. M. Budhyani, and D. G. H. Divayana, “Empowerment of CSE-UCLA Model Based on Glickman Quadrant Aided by Visual Application to Evaluate the Blended Learning Program on SMA Negeri 1 Ubud,” *Journal of Theoretical and Applied Information Technology*, Vol. 96, No. 18, pp. 6203–6219, 2018.

[3] D. Thurab-Nkhosi, “The Evaluation of A Blended Faculty Development Course using the CIPP Framework,” *International Journal of Education and Development using Information and Communication Technology*, Vol. 15, No. 1, pp. 245–254, 2019.

[4] M. R. Savoie-Roskos, S. Bevan, R. Charlton, and M. I. Graf, “Approaches to Evaluating Blended Courses,” *Journal on Empowering Teaching Excellence*, Vol. 2, No. 1, pp. 3–11, 2018.

[5] Z. C. Embi, T. K. Neo, and M. Neo, “Using Kirkpatrick’s Evaluation Model in a Multimedia-based Blended Learning Environment,” *Journal of Multimedia Information System*, Vol. 4, No. 3, pp. 115–122, 2017.

[6] M. R. Istambul, and H. Supriadi, “Evaluation of Blended Learning Implementation which is Conditioned to Optimize the Mastery of Student Knowledge and Skills,” *International Journal of Engineering & Technology*, Vol. 7, No. 4.33, pp. 195–200, 2018.

[7] N. Q. Agustina, and F. Mukhtaruddin, “The CIPP Model-Based Evaluation on Integrated English Learning (IEL) Program at Language Center,” *English Language Teaching Educational Journal*, Vol. 2, No. 1, pp. 22–31, 2019.

[8] J. S. Ngala, G. M. Fongod, T. J. Orock, B. M. Ayuk, and E.A. Njenwi, “Evaluating Distance Education Programme Using Stufflebeam CIPP Model: University of Buea Cameroon,” *Jitzi Samuel Ngala Journal of Engineering Research and Application*, Vol. 9, No. 10, pp. 1–15, 2019.

[9] Y. Siswadi, G. S. Houghty, and T. Agustina, “Implementation of the CIPP Evaluation Model in Indonesian Nursing Schools,” *Jurnal Ners*, Vol. 14, No. 3, pp. 126–131, 2019.

[10] A. Sugianto, “Individual Learning Plans Program Evaluation of Planning Education in Junior High School: Discrepancy Model,” *Journal of Physics: Conference Series*, Vol. 1422, pp. 1–6, 2020.

[11] R. S. Ambida, and R. A. Cruz, “Extent of Compliance of a Higher Education Institution for a University System,” *Science Journal of Education*, Vol. 5, No. 3, pp. 90–99, 2017.

[12] A. Ibrahim, Yusmaniarti, Y. R. Sofita, R. Sepdela, Z. E. Putra, D. T. Ananda, and M. M. Febrianti, “The Effectiveness of Instagram Features as a Sales Promotion Media Using Discrepancy Evaluation Model Method in Increasing Customer Loyalty,” *Advances in Intelligent Systems Research*, Vol. 172, pp. 665–673, 2020.

[13] K. D. I. Agustini, I. M. Suarjana, I. N. L. Jayanta, and N. T. Renda, “Tat Twam Asi Based Role-Playing Learning Model in Social Studies Knowledge Competence,” *International Journal of Elementary Education*, Vol. 4, No. 20, pp. 187–199, 2020.

[14] H. Perbowosari, “The Local Wisdom Value of Mandhasiya Tradition (Study of Hindu Education),” *International Journal of Hindu Science and Religious Studies*, Vol. 3, No.1, pp. 1–12, 2019.

[15] A. A. A. N. S. R. Gorda, “The Local Knowledge Perspective of Banking Laws,” *International Journal of Business, Economics and Law*, Vol. 14, No. 4, pp. 152–156, 2017.

[16] B. Sahin, T. L. Yip, P. H. Tseng, M. Kabak, and A. Soylu, “An Application of a Fuzzy TOPSIS Multi-Criteria Decision Analysis Algorithm for Dry Bulk Carrier Selection,” *Information*, Vol. 11, pp. 1–16, 2020.

[17] N. B. Kore, K. Ravi, and S. B. Patil, “A Simplified Description of FUZZY TOPSIS Method for Multi Criteria Decision Making," *International Research Journal of Engineering and Technology (IRJET)*, Vol. 4, No. 5, pp. 2047–2050, 2017.

[18] E. Roszkowska, and M. Filipowicz-Chomko, “Measuring Sustainable Development in the Education Area Using Multi-Criteria Methods: A Case Study,” *Central European Journal of Operations Research*, Vol. 28, No. 4, pp. 1219–1241, 2020.

[19] Y. Çelikbilek, and F. Tüysüz, “An In-Depth Review of Theory of the TOPSIS Method: An Experimental Analysis,” *Journal of Management Analytics*, Vol. 7, No. 2, pp. 281–300, 2020.

[20] K. P. Yoon, and W. K. Kim, “The Behavioral TOPSIS,” *Expert Systems with Applications*, Vol. 89, pp. 266–272, 2017.

[21] R. Rahim, S. Supiyandi, A. P. U. Siahaan, T. Listyorini, A. P. Utomo, W. A. Triyanto, Y. Irawan, S. Aisyah, M. Khairani, S. Sundari, and K. Khairunnisa, “TOPSIS Method Application for Decision Support System in Internal Control for Selecting Best Employees,” *Journal of Physics: Conference Series*, Vol. 1028, pp. 1–8, 2018.

[22] P. K. Parida, “A Multi-Attributes Decision Making Model Based on Fuzzy TOPSIS for Positive and Negative Ideal Solutions with Ranking Order,” *International Journal of Civil Engineering and Technology (IJCIET)*, Vol. 9, No. 6, pp. 190–198, 2018.

[23] C. Dziuban, C. R. Graham, P. D. Moskal, A. Norberg, and N. Sicilia, “Blended learning: the new normal and emerging technologies,” *International Journal of Educational Technology in Higher Education*, Vol. 15, No. 3, pp. 1–16, 2018.

[24] F. Harahap, N. E. A. Nasution, and B. Manurung, “The Effect of Blended Learning on Student’s Learning Achievement and Science Process Skills in Plant Tissue Culture Course,” *International Journal of Instruction*, Vol. 12, No.1, pp. 521–538, 2019.

[25] T. S. Y. Masadeh, “Blended Learning: Issues Related to Successful Implementation,” *International Journal of Scientific Research and Management (IJSRM)*, Vol. 9, No. 10, pp. 1897–1907, 2021.

[26] D. G. H. Divayana, P. W. A. Suyasa, and I. B. G. S. Abadi. “Digital library Evaluation Application Based on Combination of CSE-UCLA with Weighted Product,” *Journal of Engineering and Applied Sciences*, Vol. 14, No. 4, pp. 1318–1330, 2019.

[27] D. G. H. Divayana, P. W. A. Suyasa, and N. K. Widiartini, “An Innovative Model as Evaluation Model for Information Technology-Based Learning at ICT Vocational Schools,” *Heliyon*, Vol. 7, No. 2, pp. 1–13, 2021.

[28] Najuah, Syarifah, and R. Sidiq. “The Development and Utilization of E-Learning Media using the Edmodo Applications for Statistic Course,” *Advances in Social Science, Education and Humanities Research (ASSEHR)*, Vol. 208, pp. 123–126, 2019.

[29] D. G. H. Divayana, P. W. A. Suyasa, and I. B. G. S. Abadi. “The Effectiveness of Evaluation Application Implementation Based on Alkin(CSE-UCLA)-Weighted Product Model to Evaluate the Digital Library Services as Education Supporting Facilities,” *Journal of Physics: Conference Series*, Vol. 1402, No. 2, pp. 1–7, 2019.

[30] M. Dalimunte, and M. Salmiah, “Students’ Ability at Changing Direct into Indirect Speech and Indirect into Direct Speech,” *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, Vol. 2, No. 2, pp. 178–185, 2019.

[31] Sutirna, “Subject Teachers’ Perceptions of Academic Mentoring and Counseling Services,” *COUNS-EDU: The International Journal of Counseling and Education*, Vol. 4, No. 4, pp. 129–133, 2019.

[32] S. A. Sari, and Y. S. Rezeki, “The Development of An Ingenious Circuit Based on Chemo-Edutainment Learning,” *International Journal of Educational Research Review*, Vol. 4, No. 1, pp. 15–25, 2019.

[33] I. K. Yulina, A. Permanasari, H. Hernani, and W. Setiawan, “Analytical Thinking Skill Profile and Perception of Pre Service Chemistry Teachers in Analytical Chemistry Learning,” *Journal of Physics: Conference Series*, Vol. 1157, pp. 1–7, 2019.

[34] S. Nawawi, Nizkon, and A. T. Azhari, “Analysis of the Level of Critical Thinking Skills of Students in Biological Materials at Muhammadiyah High School in Palembang City,” *Universal Journal of Educational Research*, Vol. 8, No. 3D, pp. 47–53, 2020.

[35] R. Mantasiah, Yusri, and Jufri, “Semantic Feature Analysis Model: Linguistics Approach in Foreign Language Learning Material Development. *International Journal of Instruction*, Vol. 13, No. 1, pp. 185–196, 2020.

[36] A. A. G. Agung, I. G. P. Sudiarta, and D. G. H. Divayana, “The Quality Evaluation of School Management Model Based on Balinese Local Wisdom Using Weighted Product Calculation,” *Journal of Theoretical and Applied Information Technology*, Vol. 96, No. 19, pp. 6570–6579, 2018.

[37] D. G. H. Divayana, I. P. W. Ariawan, and A. Adiarta, “Development of Countenance Application Oriented on Combining ANEKA-Tri Hita Karana as A Mobile Web to Evaluate the Computer Knowledge and Morality,” *International Journal of Interactive Mobile Technologies (iJIM)*, Vol. 13, No. 12, pp. 81–103, 2019.

[38] L. Ma, and C. S. Lee, “Evaluating the effectiveness of blended learning using the ARCS model,” *Journal of Computer Assisted Learning*, Vol. 37, No. 5, pp. 1397–1408, 2021.

[39] L. Martín-Martínez, V. Sainz, and F. Rodríguez-Legendre, “Evaluation of a blended learning model for pre-service teachers,” *Knowledge Management & E-Learning*, Vol. 12, No.2, pp. 147–164, 2020.

[40] Sukirman, Y. Masduki, Suyono, D. Hidayati, H. C. A. Kistoro, and S. Ru’iya, “Effectiveness of Blended Learning in the New Normal Era,” *International Journal of Evaluation and Research in Education (IJERE)*, Vol. 11, No.2, pp. 628–638, 2022.

Solving the Job Shop Scheduling Problem by the Multi-Hybridization of Swarm Intelligence Techniques

Jebari Hakim[1]*, Siham Rekiek[2], Kamal Reklaoui[3]
Research Laboratory in Engineering, Innovation and Management of Industrial Systems (2ISMI), Faculty of Science and Technique of Tangier, University Abdelmalek Essaâdi, Morocco[1, 3]
The International Higher Institute of Tourism of Tangier (ISITT), Morocco[2]

***Abstract*—The industry is subject to strong competition, and customer requirements which are increasingly strong in terms of quality, cost, and deadlines. Consequently, the companies must improve their competitiveness. Scheduling is an essential tool for improving business performance. The production scheduling problem is usually an NP-hard problem, its resolution requires optimization methods dedicated to its degree of difficulty. This paper aims to develop multi-hybridization of swarm intelligence techniques to solve job shop scheduling problems. The performance of recommended techniques is evaluated by applying them to all well-known benchmark instances and comparing their results with the results of other techniques obtainable in the literature. The experiment results are concordant with other studies that have shown that the multi hybridization of swarm intelligence techniques improve the effectiveness of the method and they show how these recommended techniques affect the resolution of the job shop scheduling problem.**

Keywords*—Scheduling; Job shop; Multi-hybridization; Swarm intelligence

I. Introduction

The profitability of a manufacturing company is generally achieved by minimizing production lead times, reducing stock costs, meeting deadlines, maximizing customer satisfaction and maximizing the use of machinery. These key performance indicators are dependent on effective scheduling.

Scheduling is the process of deciding how to allocate resources among the various potential job categories. The objective is to allocate resources over a period of time to complete a series of jobs. It has been the topic of a numerous publications in the area of operational research [1-4]. This is an important decision-making process in most sectors of manufacturing and services [5]. It may also be described as a decision-making process in order to optimize goals such as achieving and reducing makespan.

Scheduling problems may be modelled as allocation issues that represent a broad class of combinatorial optimization issues. In most such cases, it is very hard to come up with the optimum solution.

Job shop scheduling problem (JSSP) is a classic problem of operational search which has been regarded as a problem of combinatorial optimization difficult since the 1950s. As for the complexity of the calculations, the JSSP is an NP-hard in the strong sense of the term [6]. As a result, even in the case of very small JSSP instances, an optimum solution cannot be ensured.

In the general, job shop scheduling problem can be formulated as follows:

- There is a set of n jobs to be treated over a set of m machines.
- A job should not go on the same machine more than once.
- There are no constraints of precedence over the operations of the various jobs.
- The operations cannot be halted.
- Each machine can handle only one job at a time.
- Every job must pass through a specific sequence of predefined operations.

Every optimization issue should have an objective function that must be kept to a minimum or maximized in order to achieve a solution. In this case, the purpose of this paper is to reduce to a minimum the total time required to complete all jobs (makespan).

For job shop scheduling problems, there are numerous reference instances for the makespan minimization case:

1) FT (3) [7]: It is a collection of three instances of the combinations (n, m) ∈ {(6, 6), (10, 10), (20, 5)}.

2) LA (40) [8]: It is a collection of forty instances of which five instances of combinations (n, m) ∈ {(10, 5), (15, 5), (20, 5), (10, 10), (15, 10), (20, 10), (30, 10), (15, 15)}.

3) ABZ (5) [9]: It is a collection of five instances of the combinations (n, m) ∈ {(10, 10), (20, 15)}.

4) ORB (10) [10]: It is a collection of ten instances of the format (n, m) = (10, 10).

5) YN (4) [11]: It is a collection of four instances of the format (n, m) = (20, 20).

6) SWV (20) [12]: It is a collection of twenty instances of the combinations (n, m) ∈ {(20, 10), (20, 15), (50, 10), (50, 10)}.

*Corresponding Author

7) TA (80) [13]: It is a collection of eighty instances with ten instances for each of the formats (n, m) ∈ {(15, 15), (20, 15), (20, 20), (30, 15), (30, 30), (50, 15), (50, 20), (100, 20)}.

8) DMU (80) [14]: It is a collection of eighty instances of which ten instances for each of the combinations (n, m) ∈ {(20, 15), (20, 20), (30, 15), (30, 20), (40, 15), (40, 20), (50, 15), (50, 20)}.

Because of the NP-hard kind of job shop scheduling problems, the development of an optimum schedule is very expensive and impracticable. Hence, numerous techniques are developed for this problem.

A few contributions that have been made to solve the job shop scheduling problem are summarized in Fig. 1.

Several disadvantages have been encountered in solving the job shop scheduling problem, therefore, it became evident that concentrating on a single metaheuristic to solve the JSSP is rather limited.

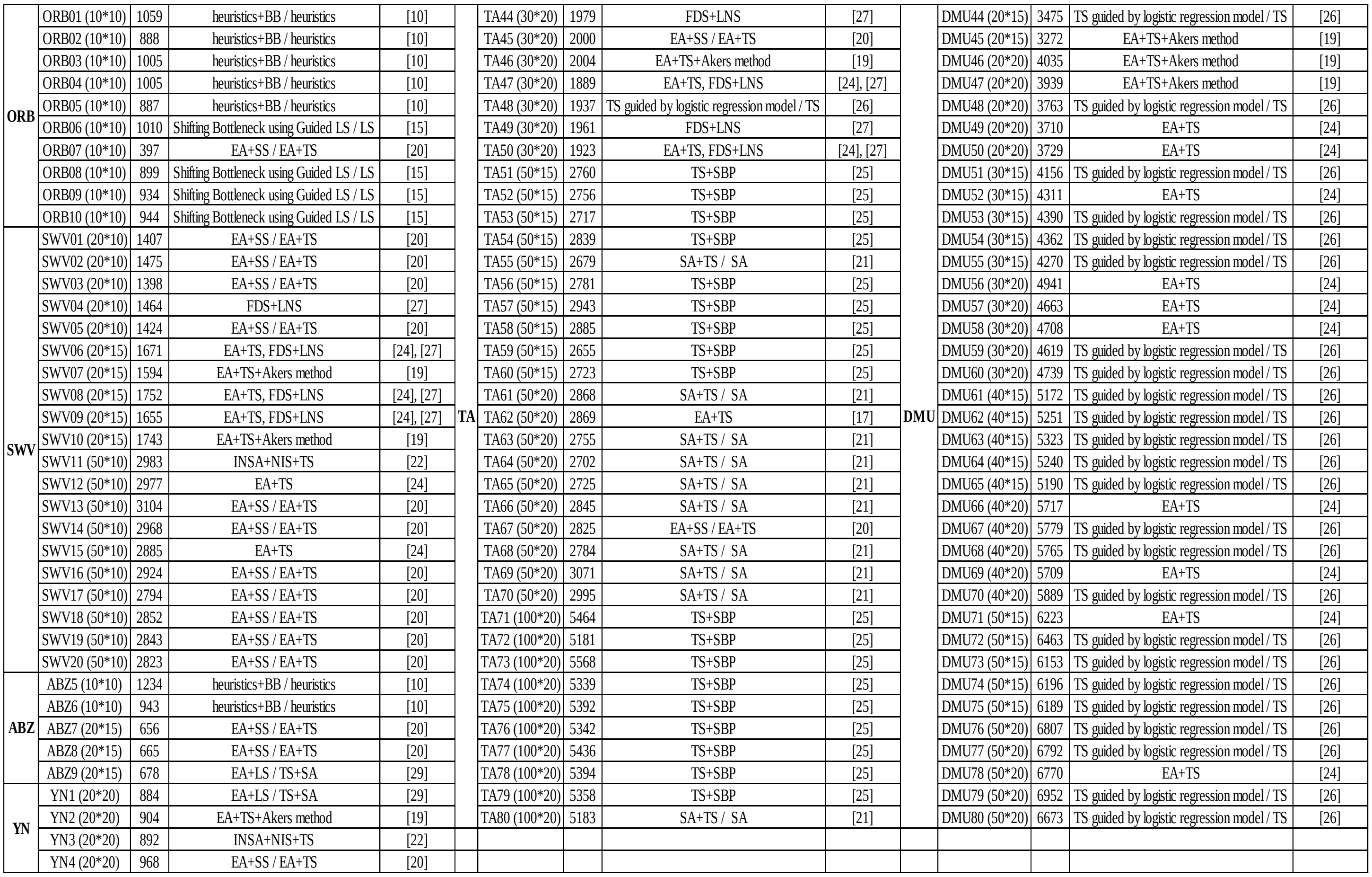
A metaheuristic is rarely as powerful in diversification as it is in intensification, the solution consists in combining a technique characterized by a high exploration capability with a technique characterized by a good exploitation of the search space.

A good balance between diversification and intensification is required to obtain the optimum performance of a hybrid optimization method, hence its efficiency and success.

However, to the authors' best knowledge, very few articles can be found in the literature which investigate the multi hybridization of swarm intelligence techniques and conduct a thorough analysis.

The remainder of the paper is organized as follows: The material and method will be described in section II. The experimental findings and discussion are discussed in Section III. The conclusion is presented under Section IV.

Instances		BKS	Resolution Method	Reference	Instances		BKS	Resolution Method	Reference	Instances		BKS	Resolution Method	Reference
FT	FT06 (6*6)	55	BB	[18]	TA	TA01 (15*15)	1231	EA+SS / EA+TS	[20]	DMU	DMU01 (20*15)	2563	EA+SS / EA+TS	[20]
	FT10 (10*10)	930	BB	[18]		TA02 (15*15)	1244	SA+TS / SA	[21]		DMU02 (20*15)	2706	EA+SS / EA+TS	[20]
	FT20 (20*5)	1165	BB	[18]		TA03 (15*15)	1218	EA+SS / EA+TS	[20]		DMU03 (20*15)	2731	EA+SS / EA+TS	[20]
LA	LA01 (10*5)	666	heuristics+BB / heuristics	[10]		TA04 (15*15)	1175	TS+SBP	[25]		DMU04 (20*15)	2669	EA+SS / EA+TS	[20]
	LA02 (10*5)	655	heuristics+BB / heuristics	[10]		TA05 (15*15)	1224	EA+SS / EA+TS	[20]		DMU05 (20*15)	2749	EA+SS / EA+TS	[20]
	LA03 (10*5)	597	heuristics+BB / heuristics	[10]		TA06 (15*15)	1238	EA+SS / EA+TS	[20]		DMU06 (20*20)	3244	SA	[23]
	LA04 (10*5)	590	heuristics+BB / heuristics	[10]		TA07 (15*15)	1227	EA+SS / EA+TS	[20]		DMU07 (20*20)	3046	SA	[23]
	LA05 (10*5)	593	heuristics+BB / heuristics	[10]		TA08 (15*15)	1217	EA+SS / EA+TS	[20]		DMU08 (20*20)	3188	SA	[23]
	LA06 (15*5)	926	heuristics+BB / heuristics	[10]		TA09 (15*15)	1274	EA+SS / EA+TS	[20]		DMU09 (20*20)	3092	EA+SS / EA+TS	[20]
	LA07 (15*5)	890	heuristics+BB / heuristics	[10]		TA10 (15*15)	1241	EA+SS / EA+TS	[20]		DMU10 (20*20)	2984	SA	[23]
	LA08 (15*5)	863	heuristics+BB / heuristics	[10]		TA11 (20*15)	1357	TS+CP	[16]		DMU11 (30*15)	3430	EA+TS	[24]
	LA09 (15*5)	951	heuristics+BB / heuristics	[10]		TA12 (20*15)	1367	EA+SS / EA+TS	[20]		DMU12 (30*15)	3492	TS guided by logistic regression model / TS	[26]
	LA10 (15*5)	958	heuristics+BB / heuristics	[10]		TA13 (20*15)	1342	EA+SS / EA+TS	[20]		DMU13 (30*15)	3681	EA+TS+Akers method	[19]
	LA11 (20*5)	1222	heuristics+BB / heuristics	[10]		TA14 (20*15)	1345	SA+TS / SA	[21]		DMU14 (30*15)	3394	EA+SS / EA+TS	[20]
	LA12 (20*5)	1039	heuristics+BB / heuristics	[10]		TA15 (20*15)	1339	SA	[23]		DMU15 (30*15)	3343	EA+SS / EA+TS	[20]
	LA13 (20*5)	1150	heuristics+BB / heuristics	[10]		TA16 (20*15)	1360	EA+SS / EA+TS	[20]		DMU16 (30*20)	3751	EA+TS+Akers method	[19]
	LA14 (20*5)	1292	heuristics+BB / heuristics	[10]		TA17 (20*15)	1462	EA+SS / EA+TS	[20]		DMU17 (30*20)	3814	TS guided by logistic regression model / TS	[26]
	LA15 (20*5)	1207	heuristics+BB / heuristics	[10]		TA18 (20*15)	1396	EA+SS / EA+TS	[20]		DMU18 (30*20)	3844	EA+TS+Akers method	[19]
	LA16 (10*10)	945	heuristics+BB / heuristics	[10]		TA19 (20*15)	1332	SA	[23]		DMU19 (30*20)	3765	TS guided by logistic regression model / TS	[26]
	LA17 (10*10)	784	heuristics+BB / heuristics	[10]		TA20 (20*15)	1348	SA	[23]		DMU20 (30*20)	3710	EA+TS	[24]
	LA18 (10*10)	848	heuristics+BB / heuristics	[10]		TA21 (20*20)	1642	TS+CP	[16]		DMU21 (40*15)	4380	EA+SS / EA+TS	[20]
	LA19 (10*10)	842	heuristics+BB / heuristics	[10]		TA22 (20*20)	1600	EA+SS / EA+TS	[20]		DMU22 (40*15)	4725	EA+SS / EA+TS	[20]
	LA20 (10*10)	902	heuristics+BB / heuristics	[10]		TA23 (20*20)	1557	EA+SS / EA+TS	[20]		DMU23 (40*15)	4668	EA+SS / EA+TS	[20]
	LA21 (15*10)	1046	EA+LS+Crossover	[28]		TA24 (20*20)	1644	FDS+LNS	[27]		DMU24 (40*15)	4648	EA+SS / EA+TS	[20]
	LA22 (15*10)	927	heuristics+BB / heuristics	[10]		TA25 (20*20)	1595	INSA+NIS+TS	[22]		DMU25 (40*15)	4164	EA+SS / EA+TS	[20]
	LA23 (15*10)	1032	heuristics+BB / heuristics	[10]		TA26 (20*20)	1643	EA+TS+Akers method	[19]		DMU26 (40*20)	4647	EA+TS+Akers method	[19]
	LA24 (15*10)	935	heuristics+BB / heuristics	[10]		TA27 (20*20)	1680	EA+SS / EA+TS	[20]		DMU27 (40*20)	4848	EA+SS / EA+TS	[20]
	LA25 (15*10)	977	heuristics+BB / heuristics	[10]		TA28 (20*20)	1603	SA	[23]		DMU28 (40*20)	4692	EA+SS / EA+TS	[20]
	LA26 (20*10)	1218	heuristics+BB / heuristics	[10]		TA29 (20*20)	1625	EA+SS / EA+TS	[20]		DMU29 (40*20)	4691	EA+SS / EA+TS	[20]
	LA27 (20*10)	1235	EA+LS+Crossover	[28]		TA30 (20*20)	1584	EA+SS / EA+TS	[20]		DMU30 (40*20)	4732	EA+SS / EA+TS	[20]
	LA28 (20*10)	1216	heuristics+BB / heuristics	[10]		TA31 (30*15)	1764	EA+SS / EA+TS	[20]		DMU31 (50*15)	5640	EA+SS / EA+TS	[20]
	LA29 (20*10)	1152	EA+SS / EA+TS	[20]		TA32 (30*15)	1774	Parallel TS	[13]		DMU32 (50*15)	5927	EA+SS / EA+TS	[20]
	LA30 (20*10)	1355	heuristics+BB / heuristics	[10]		TA33 (30*15)	1791	SA	[23]		DMU33 (50*15)	5728	EA+SS / EA+TS	[20]
	LA31 (30*10)	1784	heuristics+BB / heuristics	[10]		TA34 (30*15)	1829	EA+SS / EA+TS	[20]		DMU34 (50*15)	5385	EA+SS / EA+TS	[20]
	LA32 (30*10)	1850	heuristics+BB / heuristics	[10]		TA35 (30*15)	2007	TS+SBP	[25]		DMU35 (50*15)	5635	EA+SS / EA+TS	[20]
	LA33 (30*10)	1719	heuristics+BB / heuristics	[10]		TA36 (30*15)	1819	EA+SS / EA+TS	[20]		DMU36 (50*20)	5621	EA+SS / EA+TS	[20]
	LA34 (30*10)	1721	heuristics+BB / heuristics	[10]		TA37 (30*15)	1771	EA+TS+Akers method	[19]		DMU37 (50*20)	5851	EA+SS / EA+TS	[20]
	LA35 (30*10)	1888	heuristics+BB / heuristics	[10]		TA38 (30*15)	1673	EA+SS / EA+TS	[20]		DMU38 (50*20)	5713	EA+SS / EA+TS	[20]
	LA36 (15*15)	1268	heuristics+BB / heuristics	[10]		TA39 (30*15)	1795	EA+SS / EA+TS	[20]		DMU39 (50*20)	5747	EA+SS / EA+TS	[20]
	LA37 (15*15)	1397	heuristics+BB / heuristics	[10]		TA40 (30*15)	1669	EA+TS+Akers method	[19]		DMU40 (50*20)	5577	EA+SS / EA+TS	[20]
	LA38 (15*15)	1196	SA+TS / SA	[21]		TA41 (30*20)	2005	FDS+LNS	[27]		DMU41 (20*15)	3248	EA+TS	[24]
	LA39 (15*15)	1233	heuristics+BB / heuristics	[10]		TA42 (30*20)	1937	EA+TS+Akers method	[19]		DMU42 (20*15)	3390	EA+TS	[24]
	LA40 (15*15)	1222	heuristics+BB / heuristics	[10]		TA43 (30*20)	1846	EA+TS	[24]		DMU43 (20*15)	3441	EA+TS+Akers method	[19]



ORB	ORB01 (10*10)	1059	heuristics+BB / heuristics	[10]	TA	TA44 (30*20)	1979	FDS+LNS	[27]	DMU	DMU44 (20*15)	3475	TS guided by logistic regression model / TS	[26]
	ORB02 (10*10)	888	heuristics+BB / heuristics	[10]		TA45 (30*20)	2000	EA+SS / EA+TS	[20]		DMU45 (20*15)	3272	EA+TS+Akers method	[19]
	ORB03 (10*10)	1005	heuristics+BB / heuristics	[10]		TA46 (30*20)	2004	EA+TS+Akers method	[19]		DMU46 (20*20)	4035	EA+TS+Akers method	[19]
	ORB04 (10*10)	1005	heuristics+BB / heuristics	[10]		TA47 (30*20)	1889	EA+TS, FDS+LNS	[24], [27]		DMU47 (20*20)	3939	EA+TS+Akers method	[19]
	ORB05 (10*10)	887	heuristics+BB / heuristics	[10]		TA48 (30*20)	1937	TS guided by logistic regression model / TS	[26]		DMU48 (20*20)	3763	TS guided by logistic regression model / TS	[26]
	ORB06 (10*10)	1010	Shifting Bottleneck using Guided LS / LS	[15]		TA49 (30*20)	1961	FDS+LNS	[27]		DMU49 (20*20)	3710	EA+TS	[24]
	ORB07 (10*10)	397	EA+SS / EA+TS	[20]		TA50 (30*20)	1923	EA+TS, FDS+LNS	[24], [27]		DMU50 (20*20)	3729	EA+TS	[24]
	ORB08 (10*10)	899	Shifting Bottleneck using Guided LS / LS	[15]		TA51 (50*15)	2760	TS+SBP	[25]		DMU51 (30*15)	4156	TS guided by logistic regression model / TS	[26]
	ORB09 (10*10)	934	Shifting Bottleneck using Guided LS / LS	[15]		TA52 (50*15)	2756	TS+SBP	[25]		DMU52 (30*15)	4311	EA+TS	[24]
	ORB10 (10*10)	944	Shifting Bottleneck using Guided LS / LS	[15]		TA53 (50*15)	2717	TS+SBP	[25]		DMU53 (30*15)	4390	TS guided by logistic regression model / TS	[26]
SWV	SWV01 (20*10)	1407	EA+SS / EA+TS	[20]		TA54 (50*15)	2839	TS+SBP	[25]		DMU54 (30*15)	4362	TS guided by logistic regression model / TS	[26]
	SWV02 (20*10)	1475	EA+SS / EA+TS	[20]		TA55 (50*15)	2679	SA+TS / SA	[21]		DMU55 (30*15)	4270	TS guided by logistic regression model / TS	[26]
	SWV03 (20*10)	1398	EA+SS / EA+TS	[20]		TA56 (50*15)	2781	TS+SBP	[25]		DMU56 (30*20)	4941	EA+TS	[24]
	SWV04 (20*10)	1464	FDS+LNS	[27]		TA57 (50*15)	2943	TS+SBP	[25]		DMU57 (30*20)	4663	EA+TS	[24]
	SWV05 (20*10)	1424	EA+SS / EA+TS	[20]		TA58 (50*15)	2885	TS+SBP	[25]		DMU58 (30*20)	4708	EA+TS	[24]
	SWV06 (20*15)	1671	EA+TS, FDS+LNS	[24], [27]		TA59 (50*15)	2655	TS+SBP	[25]		DMU59 (30*20)	4619	TS guided by logistic regression model / TS	[26]
	SWV07 (20*15)	1594	EA+TS+Akers method	[19]		TA60 (50*15)	2723	TS+SBP	[25]		DMU60 (30*20)	4739	TS guided by logistic regression model / TS	[26]
	SWV08 (20*15)	1752	EA+TS, FDS+LNS	[24], [27]		TA61 (50*20)	2868	SA+TS / SA	[21]		DMU61 (40*15)	5172	TS guided by logistic regression model / TS	[26]
	SWV09 (20*15)	1655	EA+TS, FDS+LNS	[24], [27]		TA62 (50*20)	2869	EA+TS	[17]		DMU62 (40*15)	5251	TS guided by logistic regression model / TS	[26]
	SWV10 (20*15)	1743	EA+TS+Akers method	[19]		TA63 (50*20)	2755	SA+TS / SA	[21]		DMU63 (40*15)	5323	TS guided by logistic regression model / TS	[26]
	SWV11 (50*10)	2983	INSA+NIS+TS	[22]		TA64 (50*20)	2702	SA+TS / SA	[21]		DMU64 (40*15)	5240	TS guided by logistic regression model / TS	[26]
	SWV12 (50*10)	2977	EA+TS	[24]		TA65 (50*20)	2725	SA+TS / SA	[21]		DMU65 (40*15)	5190	TS guided by logistic regression model / TS	[26]
	SWV13 (50*10)	3104	EA+SS / EA+TS	[20]		TA66 (50*20)	2845	SA+TS / SA	[21]		DMU66 (40*20)	5717	EA+TS	[24]
	SWV14 (50*10)	2968	EA+SS / EA+TS	[20]		TA67 (50*20)	2825	EA+SS / EA+TS	[20]		DMU67 (40*20)	5779	TS guided by logistic regression model / TS	[26]
	SWV15 (50*10)	2885	EA+TS	[24]		TA68 (50*20)	2784	SA+TS / SA	[21]		DMU68 (40*20)	5765	TS guided by logistic regression model / TS	[26]
	SWV16 (50*10)	2924	EA+SS / EA+TS	[20]		TA69 (50*20)	3071	SA+TS / SA	[21]		DMU69 (40*20)	5709	EA+TS	[24]
	SWV17 (50*10)	2794	EA+SS / EA+TS	[20]		TA70 (50*20)	2995	SA+TS / SA	[21]		DMU70 (40*20)	5889	TS guided by logistic regression model / TS	[26]
	SWV18 (50*10)	2852	EA+SS / EA+TS	[20]		TA71 (100*20)	5464	TS+SBP	[25]		DMU71 (50*15)	6223	EA+TS	[24]
	SWV19 (50*10)	2843	EA+SS / EA+TS	[20]		TA72 (100*20)	5181	TS+SBP	[25]		DMU72 (50*15)	6463	TS guided by logistic regression model / TS	[26]
	SWV20 (50*10)	2823	EA+SS / EA+TS	[20]		TA73 (100*20)	5568	TS+SBP	[25]		DMU73 (50*15)	6153	TS guided by logistic regression model / TS	[26]
ABZ	ABZ5 (10*10)	1234	heuristics+BB / heuristics	[10]		TA74 (100*20)	5339	TS+SBP	[25]		DMU74 (50*15)	6196	TS guided by logistic regression model / TS	[26]
	ABZ6 (10*10)	943	heuristics+BB / heuristics	[10]		TA75 (100*20)	5392	TS+SBP	[25]		DMU75 (50*15)	6189	TS guided by logistic regression model / TS	[26]
	ABZ7 (20*15)	656	EA+SS / EA+TS	[20]		TA76 (100*20)	5342	TS+SBP	[25]		DMU76 (50*20)	6807	TS guided by logistic regression model / TS	[26]
	ABZ8 (20*15)	665	EA+SS / EA+TS	[20]		TA77 (100*20)	5436	TS+SBP	[25]		DMU77 (50*20)	6792	TS guided by logistic regression model / TS	[26]
	ABZ9 (20*15)	678	EA+LS / TS+SA	[29]		TA78 (100*20)	5394	TS+SBP	[25]		DMU78 (50*20)	6770	EA+TS	[24]
YN	YN1 (20*20)	884	EA+LS / TS+SA	[29]		TA79 (100*20)	5358	TS+SBP	[25]		DMU79 (50*20)	6952	TS guided by logistic regression model / TS	[26]
	YN2 (20*20)	904	EA+TS+Akers method	[19]		TA80 (100*20)	5183	SA+TS / SA	[21]		DMU80 (50*20)	6673	TS guided by logistic regression model / TS	[26]
	YN3 (20*20)	892	INSA+NIS+TS	[22]										
	YN4 (20*20)	968	EA+SS / EA+TS	[20]										

Fig. 1. The Proposed Methods for Solving Job Shop Scheduling Problem in the Literature.

II. The Material and Method

Swarm Intelligence (SI) is considered to be one of the most important research fields that is applied by various scientists for problem-solving, computation, and solution optimization [4].

Swarm Intelligence is based on natural swarm systems and is defined as the collective problem resolution abilities of social animals. [30].

Swarm Intelligence is a direct outcome of self-organization in which the interactions of lower-level components establish an overall-level dynamic structure that can be considered intelligence [31].

Self-organization is established by four elements [30]:

1) Multiple interactions [31]: Information about food sources is shared between the employed bees and the onlooker bees on the dance floor to harvest and retrieve the food.

2) Positive feedback [31]: It is essentially a set of simple rules that assist in generating the complex structure. One example of this process is the recruitment of honeybees in a promising flower field.

3) Negative feedback [31]: Minimizes the impact of positive feedback and contributes to the creation of a counterbalance mechanism.

4) Fluctuation [31]: The scouts conduct research in the environment to randomly find the food source.

Various swarm intelligence methods have been proposed in the literature: Artificial Bee Colony (ABC) [32-35], Genetic Algorithm (GA) [36], Ant Colony Optimization (ACO) [37–40], Particle Swarm Optimization (PSO) [41], Cat Swarm (CSO) [42], Artificial Immune System [43], Bacterial Foraging [44], and Glowworm Swarm Optimization [45] and many more.

Swarm intelligence approaches are successfully used to solve many real issues and have yielded excellent results in comparison with other methods.

In this paper, the authors concentrate on two of the most popular swarm intelligence techniques, namely, Artificial Bee Colony (ABC) and Ant Colony Optimization (ACO).

A. Fundamentals of Artificial Bee Colony Algorithm (ABC)

The Artificial Bee Colony (ABC) is a population-based approach suggested by Karaboga and Basturk in 2007 [46] and an evolutionary algorithm based on the intelligent behavior of honeybees looking for food (nectar) [47, 48], which works by sharing information about food sources among bees in the nest.

Each position of the food source corresponds to one solution, the bees are ranked according to how they choose the food source to use. The phase of employed bee, onlooker bee and scout bee phase are the steps used in the suggested method.

The Artificial Bee Colony (ABC) algorithm has shown that it provides optimum solutions in the continuous and discreet field [1-4] [49-52] when it is confronted with noisy and multimodal optimization issues. A full overview of the use of the ABC technique is available in [47].

Analogous to other swarm techniques, the Artificial Bee Colony (ABC) method is a repetitive process.

The artificial bee colony technique (ABC) comes up with the best solution by applying four stages [53], until an end criterion is satisfied [54], these key steps of the ABC algorithm are outlined below:

1) Initialization step: It begins with a population of randomized solutions or food sources. The value of food sources determines by many factors such as their wealth, the ability to extract energy, the closeness of the nest and the amount of energy that's going to be collected from that food source.

2) Employed bee step: Every bee employed is given a source of food that it currently uses or operates. They bring together information about that particular source, their distance and direction from the nest, the cost-effectiveness of the source and they're sharing that information with some degree of probability. The number of employed honeybees will correspond to the number of food sources surrounding their beehives.

Unemployed bees wait in the nest and are seeking a source of food to explore, there are two kinds of bees unemployed:

a) Onlooker bee step: The Curious bees or Onlooker bees must look at the dance of the employed bees and then develop a source of food via information shared by the employed bees.

b) Scout bee step: Explorer or Scout bees are the ones whose source of food is abandoned and are seeking new food sources in the nesting environment. Once an adequate food source is found, scout bees become employed bees.

Register the best food source reached up to now.

From this algorithm, a colony of artificial bees (ABC) is established.

The Artificial Bee Colony (ABC) is also the same as other algorithms because it has pros and cons [55]:

1) The ABC method has a force in both local and global searches and it has been implemented with a number of optimization issues.

2) However, it has a number of parameters that need to be adjusted, randomly initialized and its local search are probabilistic.

B. Fundamentals of Ant Colony Optimization (ACO)

The Ant Colony Optimization technique (ACO) is another swarm intelligence approach based on ants' behavior looking for a food source from their nest with no visual information and using shortest pathways [56].

ACO is an optimization technique based on the population that Dorigo developed in 1992 [57] and has been successfully implemented to resolve various NP-hard combinatorial optimization issues that require to provide approximate solutions to the defined issue [58]. Algorithmically, the process of evaporating pheromones is helpful to prevent convergence towards a local optimal solution.

The steps involved in obtaining the best solution using the ant colony optimization technique are listed below [38]:

1) Building the solution space is made up of potential solutions with the help of the pheromone model.

2) These potential solutions are employed to update pheromone values in a manner that is considered to skew future sampling towards high quality solutions [59].

Ant behavior results in a self-reinforcement process: A pathway created by ants using the high level of pheromone is more followed by ants than the pathway with the low level of pheromone.

Numerous variations of ACO algorithms have been implemented in the literature [60].

C. The Proposed Methods

To enhance the original algorithms based on Swarm Intelligence, the researchers have generally hybridized them with other metaheuristics.

The Artificial Bee Colony (ABC) algorithm has the capacity to emerge from a local minimum and has a great ability to explore the global optimum which it is not immediately used, because the artificial bee colony (ABC) stocks it at every iteration.

The artificial bee colony (ABC) has been hybridized to increase its yields and effectiveness by balancing exploration and exploitation processes.

Swarm Intelligence's algorithms effectiveness is driven by two processes: exploration and exploitation [61]:

1) The exploration process: Allows the exploration of the search area in a more efficient way, and it can generate solutions that are sufficiently diverse.

2) The exploitation process: Utilizes all the information gathered from a defined problem to assist in finding new solutions that are better than the existing solutions.

An extensive exploitation and an insufficient exploration means that the system can have converge more quickly, but the likelihood of reaching the effective global optimum may be low. In addition, under-exploitation and over-exploration may result in very slow search paths to converge. Therefore, the balance between exploration and exploitation processes is crucial for achieving the optimum performance of the Swarm Intelligence method. In the literature, no non-hybridized

Swarm Intelligence algorithm is able to achieve this optimum equilibrium.

The multi-hybridization of ABC technique with ACO technique is proposed in this paper for getting a more powerful method that balances both processes, exploration and exploitation, harnesses and combines the benefits of Swarm Intelligence algorithms. The synchronous parallel hybridization [62] is applied in the proposed approach.

This approach, denoted as HABCACO involves integrating the ant colony optimization technique (ACO) into the employed bees step and/or in the onlooker bees step and/or in the scout bees step.

The HABCACO approach is described in Fig. 2.

HABCACO		
ABC	*Initialization step*	Uniform Random Generation
	Employed bee step	Add or do not add ACO method
	Onlooker bee step	Add or do not add ACO method
	Scout bee step	Add or do not add ACO method

Fig. 2. The HABCACO Approach.

Table I shown the configuration of the HABCACO techniques.

TABLE I. THE CONFIGURATION OF THE HABCACO TECHNIQUES

	ABC		
Hybrid ABC + ACO	***Employed bee phase***	***Onlooker bee phase***	***Scout bee phase***
HABCACO1	ACO		
HABCACO2	ACO	ACO	
HABCACO3	ACO	ACO	ACO
HABCACO4		ACO	
HABCACO5		ACO	ACO
HABCACO6			ACO
HABCACO7	ACO		ACO

The suggested approaches HABCACO has ABC technique as the primary algorithm that in its flow will call ACO technique for enhancement. ACO method refines the solution generated by the steps of ABC method (employed bees step and/or onlooker bees step and/or scout bees step) and produces a better solution to be used in the process of HABCACO.

The population of candidate solutions for these new approaches is initialized using the uniform random generation between a lower bound LB and an upper bound UB [63], this means that the potential solution frequently takes the form:

$$\mathrm{CS} = \mathrm{LB} + \alpha\,(\mathrm{UB} - \mathrm{LB}), \text{where } \alpha \in [0,1] \qquad (1)$$

The means to generate the candidate solution population affect an algorithm's efficiency.

Hence, the stepwise implementation of HABCACO2 and HABCACO7 are explained respectively in Fig. 3 and Fig. 4.

Initialization step:
Initialize input parameters.
Initialize the number of generations necessary for the termination criterion.
Initialize the number of solutions.
Initialize the employed bees' number (equal to solutions number)
Initialize the onlooker bees' number (equal to solutions number)
Initialize the number of scout bees
Initialize the possible solution populations by the uniform random generation.
Calculate the fitness value for each solution as follows:

$$fit_i = \frac{1}{(1+f_i)} \; if\ f_i \ \geq 0\ ;\ fit_i = 1 + \left|(f_i)\right| \; if\ f_i < 0$$

Where f_i is the value of the objective function of i^th solutions.
Identify the best solution.

Employed bee step:
For each employed bee
Create a new solution based on the function:

$$y_{ij} = x_{ij} + \emptyset_{ij}\left(x_{ij} + x_{kj}\right), k \neq i\ , i = \{1, 2\ \ldots, SN\},$$
$$j = \{1, 2\ \ldots, D\}\ , \emptyset_{ij} = Rand\ [-1, 1]$$

x_{min} , x_{max} are respectively the lower and upper boundaries of the search perimeter and y_{ij} is a new realizable dimension value of the solutions which is changed from the value of its previous solutions x_{ij}.
Apply ACO (new solution).
Determine the fitness value of these best solutions found by ACO.
Calculate the most appropriate solution.

Onlooker bee step:
For each onlooker bee
Create a new solution based on the function:

$$y_{ij} = x_{ij} + \emptyset_{ij}\left(x_{ij} + x_{kj}\right), k \neq i\ , i = \{1, 2\ \ldots, SN\},$$
$$j = \{1, 2\ \ldots, D\}\ , \emptyset_{ij} = Rand\ [-1, 1]$$

Choose the i^th solution associated with the probability value (p_i):

$$p_i = \frac{fit_i}{\sum_{k=1}^{SN} fit_k}$$

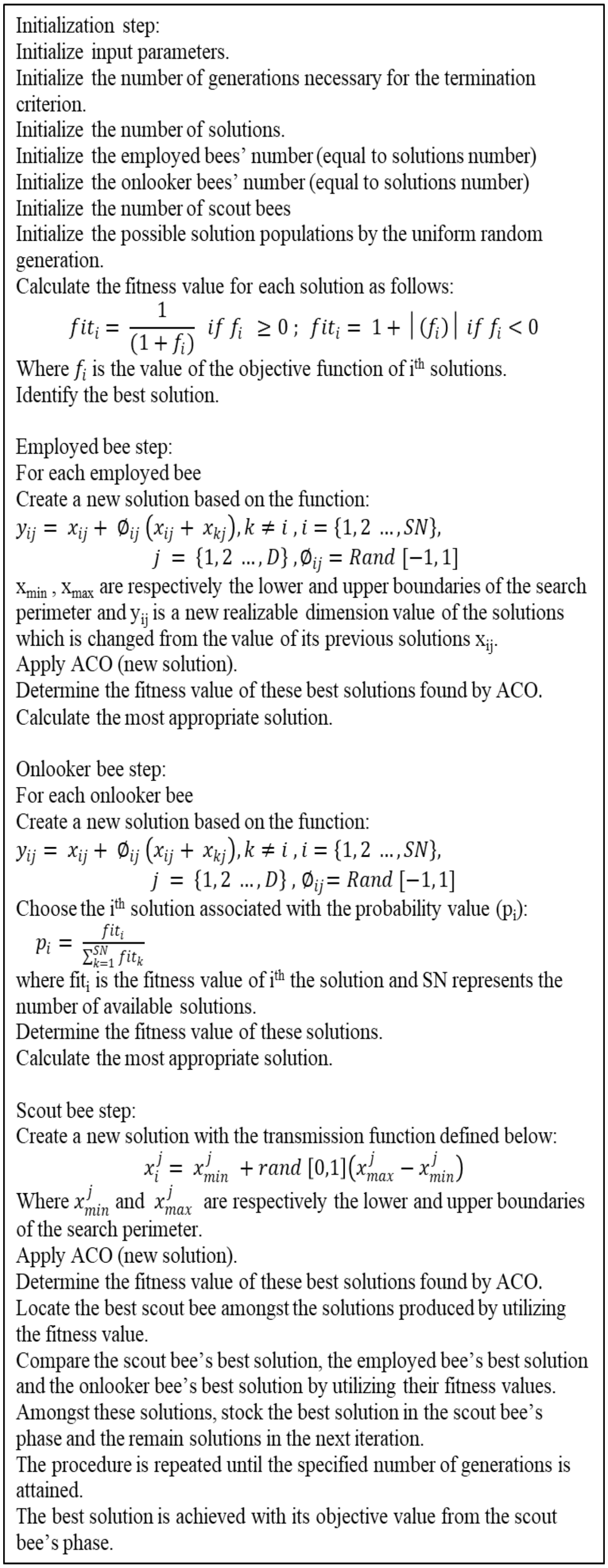
where fit_i is the fitness value of i^th the solution and SN represents the number of available solutions.
Apply ACO (new solution).
Determine the fitness value of these best solutions found by ACO.
Calculate the most appropriate solution.

Scout bee step:
Create a new solution with the transmission function defined below:

$$x_i^j = x_{min}^j + rand\ [0,1]\left(x_{max}^j - x_{min}^j\right)$$

Where x_{min}^j and x_{max}^j are respectively the lower and upper boundaries of the search perimeter.
Determine the fitness value of these solutions.
Locate the best scout bee amongst the solutions produced by utilizing the fitness value.
Compare the scout bee's best solution, the employed bee's best solution and the onlooker bee's best solution by utilizing their fitness values.
Amongst these solutions, stock the best solution in the scout bee's phase and the remain solutions in the next iteration.
The procedure is repeated until the specified number of generations is attained.
The best solution is achieved with its objective value from the scout bee's phase.

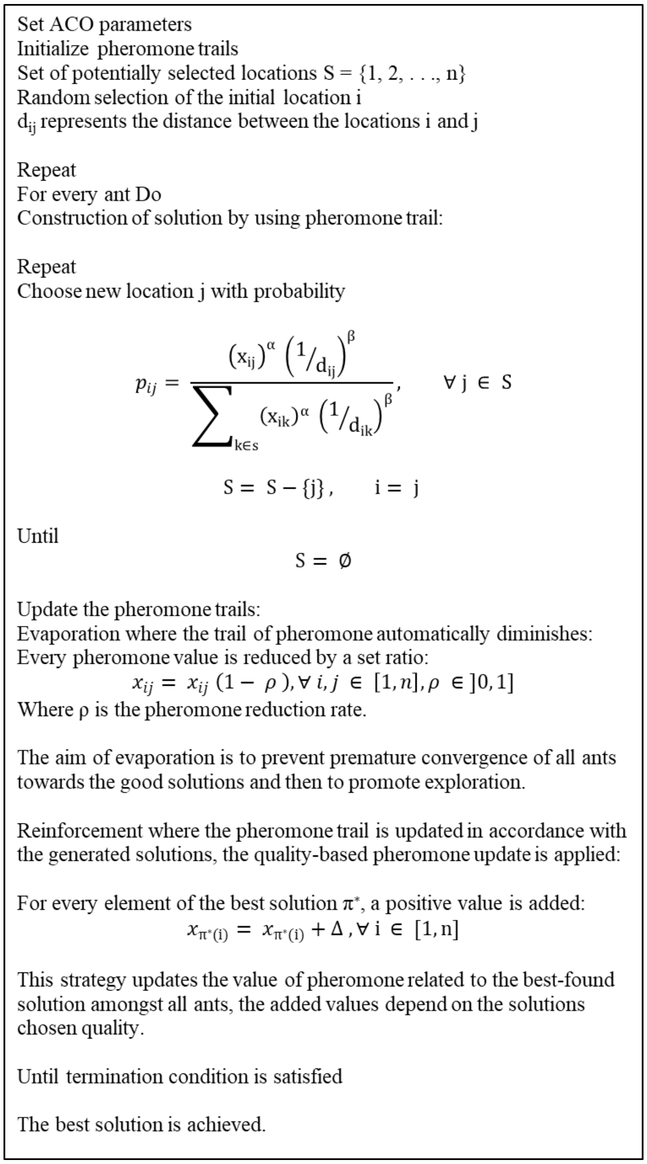
Fig. 3. The Procedure of HABCACO2.



```
Initialization step:
Initialize input parameters.
Initialize the number of generations necessary for the termination criterion.
Initialize the number of solutions.
Initialize the employed bees' number (equal to solutions number)
Initialize the onlooker bees' number (equal to solutions number)
Initialize the number of scout bees
Initialize the possible solution populations by the uniform random generation.
Calculate the fitness value for each solution as follows:
    $fit_i = \frac{1}{(1+f_i)}\ if\ f_i \geq 0\ ;\ fit_i = 1 + |(f_i)|\ if\ f_i < 0$
Where $f_i$ is the value of the objective function of i^th solutions.
Identify the best solution.

Employed bee step:
For each employed bee
Create a new solution based on the function:
$y_{ij} = x_{ij} + \emptyset_{ij}\left(x_{ij} + x_{kj}\right), k \neq i, i = \{1, 2\ \ldots, SN\},$
    $j = \{1, 2\ \ldots, D\}, \emptyset_{ij} = Rand\ [-1, 1]$
$x_{min}$, $x_{max}$ are respectively the lower and upper boundaries of the search perimeter and $y_{ij}$ is a new realizable dimension value of the solutions which is changed from the value of its previous solutions $x_{ij}$.
Apply ACO (new solution).
Determine the fitness value of these best solutions found by ACO.
Calculate the most appropriate solution.

Onlooker bee step:
For each onlooker bee
Create a new solution based on the function:
$y_{ij} = x_{ij} + \emptyset_{ij}\left(x_{ij} + x_{kj}\right), k \neq i, i = \{1, 2\ \ldots, SN\},$
    $j = \{1, 2\ \ldots, D\}, \emptyset_{ij} = Rand\ [-1, 1]$
Choose the i^th solution associated with the probability value ($p_i$):
    $p_i = \frac{fit_i}{\sum_{k=1}^{SN} fit_k}$
where $fit_i$ is the fitness value of i^th the solution and SN represents the number of available solutions.
Determine the fitness value of these solutions.
Calculate the most appropriate solution.

Scout bee step:
Create a new solution with the transmission function defined below:
    $x_i^j = x_{min}^j + rand\ [0,1]\left(x_{max}^j - x_{min}^j\right)$
Where $x_{min}^j$ and $x_{max}^j$ are respectively the lower and upper boundaries of the search perimeter.
Apply ACO (new solution).
Determine the fitness value of these best solutions found by ACO.
Locate the best scout bee amongst the solutions produced by utilizing the fitness value.
Compare the scout bee's best solution, the employed bee's best solution and the onlooker bee's best solution by utilizing their fitness values.
Amongst these solutions, stock the best solution in the scout bee's phase and the remain solutions in the next iteration.
The procedure is repeated until the specified number of generations is attained.
The best solution is achieved with its objective value from the scout bee's phase.
```

Fig. 4. The Procedure of HABCACO7.

The ant colony optimization technique (ACO) is primarily made up of two phases: construction of the solution and updating of pheromone. The algorithm of ant colony optimization (ACO) technique is described in Fig. 5:



```
Set ACO parameters
Initialize pheromone trails
Set of potentially selected locations S = {1, 2, . . ., n}
Random selection of the initial location i
$d_{ij}$ represents the distance between the locations i and j

Repeat
For every ant Do
Construction of solution by using pheromone trail:

Repeat
Choose new location j with probability

    $p_{ij} = \frac{(x_{ij})^{\alpha}\left({}^{1}/_{d_{ij}}\right)^{\beta}}{\sum_{k \in s}(x_{ik})^{\alpha}\left({}^{1}/_{d_{ik}}\right)^{\beta}}, \quad \forall j \in S$

    $S = S - \{j\}, \quad i = j$

Until
    $S = \emptyset$

Update the pheromone trails:
Evaporation where the trail of pheromone automatically diminishes:
Every pheromone value is reduced by a set ratio:
    $x_{ij} = x_{ij}\ (1 - \rho), \forall\ i, j \in [1, n], \rho \in ]0, 1]$
Where ρ is the pheromone reduction rate.

The aim of evaporation is to prevent premature convergence of all ants towards the good solutions and then to promote exploration.

Reinforcement where the pheromone trail is updated in accordance with the generated solutions, the quality-based pheromone update is applied:

For every element of the best solution $\pi^*$, a positive value is added:
    $x_{\pi^*(i)} = x_{\pi^*(i)} + \Delta, \forall\ i \in [1, n]$

This strategy updates the value of pheromone related to the best-found solution amongst all ants, the added values depend on the solutions chosen quality.

Until termination condition is satisfied

The best solution is achieved.
```

Fig. 5. The Procedure of ACO.

III. Results and Discussion

In order to justify the effectiveness of the proposed approaches HABCACO, they were simulated across a set of 250 benchmark instances from the job shop scheduling literature: FT [7], LA [8], ABZ [9], ORB [10], YN [11], SWV [12], TA [13], DMU [14], CAR [18], and compared with the best-known solution (BKS) obtained through other Techniques.

The performance of the HABCACO methods was also compared to two other advanced techniques available in the literature: HABCGA [2] and HABCPSOGA [3] to demonstrate the efficiency of HABCACO techniques in resolving job shop scheduling issues.

The HABCACO approaches have been implemented within Java and all calculation experiments were carried out on an Intel Core i7 computer with a speed of 2.5 GHz and 8 GB RAM memory under Windows 10.

All simulation and test processes were performed with the same configuration settings.

The results of the benchmark instances simulation are summarized in Table II.

The calculation results demonstrate that only the proposed HABCACO3 technique produced 100% of the best-known solution in all benchmark instances: FT (3), LA (40), ORB (10), SWV (20), ABZ (5), YN (4), TA (80), DMU (80) and CAR (8).

As shown in Table II:

1) All the suggested techniques HABCACO have given results that are 100% equal to the best-known solutions in FT (3), LA (40), ORB (10), SWV (20), ABZ (5), YN (4), and CAR (8).

2) Only the suggested techniques HABCACO3 and HABCACO7 produced 100% of the results equal to the best-known solutions in TA (80).

3) Only the suggested technique HABCACO3 produced 100% of the results equal to the best-known solutions in DMU (80).

The classification of the suggested HABCACO methods in terms of performance according to the hybridization number is illustrated in Table III.

From Table III, it can be concluded that:

1) The suggested HABCACO techniques provided the best results in comparison to the results achieved through other methods.

2) The suggested technique HABCACO hybridized in its three steps provided the best result.

3) The suggested techniques HABCACO hybridized in its two steps provided better results in comparison to the results achieved through the suggested techniques HABCACO hybridized in only one step.

4) The suggested technique HABCACO hybridized in its two steps (employed bees step and scout bees step) provided better results in comparison to the results achieved through the suggested techniques HABCACO hybridized in its two steps (onlooker bees step and scout bees step).

5) The suggested technique HABCACO hybridized in its two steps (employed bees step and scout bees step) provided better results in comparison to the results achieved through the suggested techniques HABCACO hybridized in its two steps (employed bees step and onlooker bees step).

6) The suggested technique HABCACO hybridized in its scout bees step provided better results in comparison to the results achieved through the suggested techniques HABCACO hybridized in its employed bees step or onlooker bees step.

7) The suggested technique HABCACO hybridized in its scout bees step provided better results in comparison with the results achieved through the suggested techniques HABCACO hybridized in its onlooker bees step.

The figures showed that the suggested techniques surpassed other methods with regard to the total number of benchmark instances successfully resolved and the quality of the solutions.

In order to confirm also the performance of the HABCACO technique in resolving job shop scheduling problems, it was compared to the HABCGA and HABCPSOGA methods available in the literature.

TABLE II. THE RESULTS OF BENCHMARK INSTANCES SIMULATION

HABCACO	**Benchmark Instances**																		**Total Number of Benchmark Instances**	
	FT (3)		***LA (40)***		***ORB (10)***		***SWV (20)***		***ABZ (5)***		***YN (4)***		***TA (80)***		***DMU (80)***		***CAR (8)***		***250***	
HABCACO1	3	100,00%	40	100,00%	10	100,00%	20	100,00%	5	100,00%	4	100,00%	73	91,25%	74	92,50%	8	100,00%	237	94,80%
HABCACO2	3	100,00%	40	100,00%	10	100,00%	20	100,00%	5	100,00%	4	100,00%	77	96,25%	78	97,50%	8	100,00%	245	98,00%
HABCACO3	3	100,00%	40	100,00%	10	100,00%	20	100,00%	5	100,00%	4	100,00%	80	100,00%	80	100,00%	8	100,00%	250	100,00%
HABCACO4	3	100,00%	40	100,00%	10	100,00%	20	100,00%	5	100,00%	4	100,00%	74	92,50%	76	95,00%	8	100,00%	240	96,00%
HABCACO5	3	100,00%	40	100,00%	10	100,00%	20	100,00%	5	100,00%	4	100,00%	77	96,25%	79	98,75%	8	100,00%	246	98,40%
HABCACO6	3	100,00%	40	100,00%	10	100,00%	20	100,00%	5	100,00%	4	100,00%	76	95,00%	76	95,00%	8	100,00%	242	96,80%
HABCACO7	3	100,00%	40	100,00%	10	100,00%	20	100,00%	5	100,00%	4	100,00%	80	100,00%	78	97,50%	8	100,00%	248	99,20%

TABLE III. THE RANKING OF THE PROPOSED TECHNIQUES HABCACO

Ranking	Hybridation number	HABCACO	ABC		
			Employed bee step	*Onlooker bee step*	*Scout bee step*
7	1	HABCACO1	ACO		
4	2	HABCACO2	ACO	ACO	
1	3	HABCACO3	ACO	ACO	ACO
6	1	HABCACO4		ACO	
3	2	HABCACO5		ACO	ACO
5	1	HABCACO6			ACO
2	2	HABCACO7	ACO		ACO

As shown in Table IV:

1) The suggested technique HABCACO hybridized in its three phases with ACO method provided similar result in comparison to the results achieved through HABCGA methods hybridized in its three phases with GA method.

2) The suggested technique HABCACO hybridized in its three steps with ACO method provided better result in comparison to the results achieved through HABCGA methods hybridized in its two steps with GA method.

3) The suggested technique HABCACO hybridized in its three steps with ACO method provided better result in comparison to the results achieved through HABCGA methods hybridized in its one step with GA method.

4) The suggested technique HABCACO hybridized in its two steps with ACO method (employed bee step and scout bee step) provided better result in comparison to the results achieved through HABCGA methods hybridized in its two steps (employed bee step and scout bee step) with GA method.

5) The suggested technique HABCACO hybridized in its two steps with ACO method (onlooker bee step and scout bee step) provided better result in comparison with the results achieved through HABCGA methods hybridized in its two steps (onlooker bee step and scout bee step) with GA method.

6) The suggested technique HABCACO hybridized in it scout bee step with ACO method provided better result in comparison to the results achieved through HABCGA methods hybridized in its scout bee step with GA method.

7) The suggested technique HABCACO hybridized in it onlooker bee step with ACO method provided better result in comparison with the results achieved through HABCGA methods hybridized in its onlooker bee step with GA method.

8) The HABCGA methods hybridized in its two steps with GA method (employed bee step and onlooker bee step) provided better result in comparison to the results achieved through the suggested technique HABCACO hybridized in its two steps (employed bee step and onlooker bee step) with ACO method.

9) The HABCGA methods hybridized in its employed bee steps with GA method provided better result in comparison to the results achieved through the suggested technique HABCACO hybridized in employed bee step with ACO method.

The classification of the suggested methods HABCACO and the HABCGA methods in terms of performance according to the hybridization number and the algorithm type hybridized is tabulated in Table V.

TABLE IV. THE PERFORMANCE COMPARISON OF PROPOSED METHODS HABCACO WITH THE OTHER OPTIMIZATION ALGORITHMS HABCGA

	Hybrid ABC (ACO)						Hybrid ABC (GA)					
Hybridation number	HABCACO	ABC				Ranking	HABCGA	ABC				Ranking
		Employed bee step	*Onlooker bee step*	*Scout bee step*				*Employed bee step*	*Onlooker bee step*	*Scout bee step*		
1	HABCACO1	ACO			94,80%	7	HABCGA1	GA			95,48%	5
2	HABCACO2	ACO	ACO		98,00%	4	HABCGA2	GA	GA		99,80%	2
3	HABCACO3	ACO	ACO	ACO	100,00%	1	HABCGA3	GA	GA	GA	100,00%	1
1	HABCACO4		ACO		96,00%	6	HABCGA4		GA		94,03%	7
2	HABCACO5		ACO	ACO	98,40%	3	HABCGA5		GA	GA	96,93%	4
1	HABCACO6			ACO	96,80%	5	HABCGA6			GA	95,35%	6
2	HABCACO7	ACO		ACO	99,20%	2	HABCGA7	GA		GA	98,40%	3

TABLE V. THE RANKING OF THE PROPOSED TECHNIQUES HABCACO AND HABCGA

Ranking	**Hybrid ABC (ACO) / Hybrid ABC (GA)**	**ABC**				**Hybridation number**
		Employed bee step	***Onlooker bee step***	***Scout bee step***		
1	HABCACO3	ACO	ACO	ACO	100,00%	3
1	HABCGA3	GA	GA	GA	100,00%	3
2	HABCGA2	GA	GA		99,80 %	2
3	HABCACO7	ACO		ACO	99,20 %	2
4	HABCACO5		ACO	ACO	98,40 %	2
4	HABCGA7	GA		GA	98,40 %	2
5	HABCACO2	ACO	ACO		98,00 %	2
6	HABCGA5		GA	GA	96,93 %	2
7	HABCACO6			ACO	96,80 %	1
8	HABCACO4		ACO		96,00 %	1
9	HABCGA1	GA			95,48 %	1
10	HABCGA6			GA	95,35 %	1
11	HABCACO1	ACO			94,80 %	1
12	HABCGA4		GA		94,03 %	1

From Table VI it can be concluded that:

1) The suggested technique HABCACO hybridized in its three steps with ACO method provided similar result in comparison to the results achieved through HABCPSOGA methods hybridized in its three steps with GA and PSO methods.

2) The suggested technique HABCACO hybridized in its three steps with ACO method provided better result in comparison to the results achieved through HABCPSOGA methods hybridized in its two steps with GA and PSO methods.

3) The suggested technique HABCACO hybridized in its two steps with ACO method (employed bee step and scout bee step) provided equal result in comparison to the results achieved through HABCPSOGA methods hybridized in its two steps (employed bee step by GA method and scout bee step PSO method).

4) The suggested technique HABCACO hybridized in its two steps with ACO method (employed bee step and scout bee step) provided better result in comparison to the results achieved through HABCPSOGA methods hybridized in its two steps (employed bee step by PSO method and scout bee step GA method).

5) The suggested technique HABCACO hybridized in its two steps with ACO method (onlooker bee step and scout bee step) provided better result in comparison to the results achieved through HABCPSOGA methods hybridized in its two steps (onlooker bee step by GA method and scout bee step by PSO method).

6) The suggested technique HABCACO hybridized in its two steps with ACO method (onlooker bee step and scout bee step) provided better result in comparison to the results achieved through HABCPSOGA methods hybridized in its two steps (onlooker bee step by PSO method and scout bee step by GA method).

7) The suggested technique HABCACO hybridized in its two steps with ACO method (employed bee step and onlooker bee step) provided better result in comparison to the results achieved through the suggested methods HABCPSOGA hybridized in its two steps (employed bee step by GA method and onlooker bee step by PSO method).

TABLE VI. THE PERFORMANCE COMPARISON OF PROPOSED METHODS HABCACO WITH THE OTHER OPTIMIZATION ALGORITHMS HABCPSOGA

	Hybrid ABC (ACO)						**Hybrid ABC (GA // PSO)**						
Hybridation number	**HABCACO**	**ABC**				**Ranking**	**HABCPSOGA**	**ABC**				**Ranking**	
		Employed bee step	***Onlooker bee step***	***Scout bee step***				***Employed bee step***	***Onlooker bee step***	***Scout bee step***			
2	HABCACO2	ACO	ACO		98,00%	4	HABCPSOGA4	GA	PSO		97,50%	7	
2	HABCACO2	ACO	ACO		98,00%	4	HABCPSOGA7	PSO	GA		98,48%	3	
3	HABCACO3	ACO	ACO	ACO	100,00%	1	HABCPSOGA2	GA	GA	PSO	100,00%	1	
3	HABCACO3	ACO	ACO	ACO	100,00%	1	HABCPSOGA6	GA	PSO	GA	100,00%	1	
3	HABCACO3	ACO	ACO	ACO	100,00%	1	HABCPSOGA9	PSO	GA	GA	100,00%	1	
2	HABCACO5		ACO	ACO	98,40%	3	HABCPSOGA1		GA	PSO	97,90%	5	
2	HABCACO5		ACO	ACO	98,40%	3	HABCPSOGA5		PSO	GA	97,85%	6	
2	HABCACO7	ACO		ACO	99,20%	2	HABCPSOGA3	GA		PSO	99,20%	2	
2	HABCACO7	ACO		ACO	99,20%	2	HABCPSOGA8	PSO		GA	98,40%	4	

8) The HABCPSOGA methods hybridized in its two steps (employed bee step by GA method and onlooker bee step by PSO method) provided better result in comparison to the results achieved through the suggested technique HABCACO hybridized in its two steps with ACO method (employed bee step and onlooker bee step).

The classification of the suggested methods HABCACO and the HABCPSOGA methods in terms of performance according to the hybridization number and the algorithm type hybridized is tabulated in Table VII.

As a result, it clearly demonstrated that HABCACO has the best performance compared to other methods HABCGA and HABCPSOGA.

The proposed approaches HABCACO are robust techniques that have the potential to solve job shop scheduling problems.

One of the main limitations of this research is that it only addresses the minimization of the total time required to perform all the jobs (makespan).

TABLE VII. THE RANKING OF THE PROPOSED TECHNIQUES HABCACO AND HABCPSOGA

Ranking	**Hybrid ABC (ACO) / Hybrid ABC (GA // PSO)**	**ABC**				**Hybridation number**
		Employed bee step	***Onlooker bee step***	***Scout bee step***		
1	HABCACO3	ACO	ACO	ACO	100,00%	3
1	HABCPSOGA2	GA	GA	PSO	100,00%	3
1	HABCPSOGA6	GA	PSO	GA	100,00%	3
1	HABCPSOGA9	PSO	GA	GA	100,00%	3
2	HABCACO7	ACO		ACO	99,20%	2
2	HABCPSOGA3	GA		PSO	99,20%	2
3	HABCPSOGA7	PSO	GA		98,48%	2
4	HABCACO5		ACO	ACO	98,40%	2
4	HABCPSOGA8	PSO		GA	98,40%	2
5	HABCACO2	ACO	ACO		98,00%	2
6	HABCPSOGA1		GA	PSO	97,90%	2
7	HABCPSOGA5		PSO	GA	97,85%	2
8	HABCPSOGA4	GA	PSO		97,50%	2

IV. CONCLUSION

Because of the high level of complexity of the job shop scheduling problems, powerful and hybrid approaches are essential to address these challenging NP issues.

A robust swarm intelligence multi-hybridization technique is the key to achieving maximum efficiency in the resolution of job shop scheduling issues.

In this article, the authors develop novel multi-hybridization approaches of swarm intelligence methods called HABCACO through hybridization of artificial bee colony (ABC) algorithm and ant colony optimization (ACO) technique in various ways to provide optimal or near optimal solutions to job shop scheduling problems.

In this new approach, HABCACO adjusts the standard artificial bee colony techniques (ABC) to balance the impact of exploration and exploitation processes in algorithm performance.

Balanced exploration and exploitation capacities can improve method performance in terms of solutions quality.

The assessment of HABCACO's performance was analyzed on 250 well-known benchmark instances of the classical OR-library of job shop scheduling problems.

The overall experimental findings clearly demonstrated that the proposed new technique surpassed other compared optimization algorithms in terms of the total number of successfully resolved benchmark instances and the global optimum attainment.

Furthermore, the experimental results clearly demonstrated that the approaches are robust, effective and reliable for solving job shop scheduling problems.

More importantly, the results showed that the suggested techniques HABCACO produced the best results in comparison to other optimization methods HABCGA and HABCPSOGA available in the literature.

Therefore, the suggested approaches are solid techniques which have the potential to solve the scheduling problem and can be applied to solve complex optimization issues.

As future research, the authors intend to apply these approaches developed in this article to another type of scheduling problems and complex optimization problems.

REFERENCES

[1] H. Jebari, S. Rekiek, S. R. Elazzouzi, and H. Samadi, “Performance comparison of three hybridization categories to solve multi-objective flow shop scheduling problem: A case study from the automotive industry,” International Journal of Advanced Computer Science and Applications, vol. 12, no. 4, pp. 680–689, 2021.

[2] H. Jebari, S. R. Elazzouzi, H. Samadi, and S. Rekiek, “The search of balance between diversification and intensification in artificial bee colony to solve job shop scheduling problem,” Journal of Theoretical and Applied Information Technology, vol. 97, no. 2, pp. 658–673, 2019.

[3] H. Jebari, S. R. Elazzouzi, H. Samadi, and S. Rekiek, “Multi hybridization of swarm intelligence methods to solve job shop scheduling problem,” Journal of Theoretical and Applied Information Technology, vol. 97, no. 16, pp. 4366–4386, 2019.

[4] H. Jebari, S. R. Elazzouzi, and H. Samadi, "The hybrid genetic algorithm for solving scheduling problems in a flexible production system," International Journal of Computer Applications, vol. 110, no. 12, pp. 22–29, 2015.

[5] B. S. Girish and N. Jawahar, "A scheduling algorithm for flexible job shop scheduling problem," 5th Annual IEEE Conference on Automation Science and Engineering, Bangalore, India, pp. 22–25, 2009.

[6] M. R. Garey and D. S. Johnson, "A Guide to the Theory of NP-completeness," Computers and Intractability, Freeman, 1979.

[7] H. Fisher, and G. L. Thompson, "Industrial Scheduling," Englewood Cliffs, NJ: Prentice-Hall, 1963.

[8] S. Lawrence, "Resource constrained project scheduling: an experimental investigation of heuristic scheduling techniques," Pittsburgh: Graduate School of Industrial Administration, 1984.

[9] J. Adams, E. Balas, and D. Zawack, "The shifting bottleneck procedure for job-shop scheduling," Management Science, vol. 34, no. 3, pp. 391–401, 1988.

[10] D. Applegate and W. Cook, "A computational study of the job-shop scheduling problem," ORSA Journal on Computing, vol. 3, no. 2, pp. 149–156, 1991.

[11] T. Yamada, and R. Nakano, "A genetic algorithm applicable to large-scale job-shop problems", Proceedings of the second International Workshop on Parallel Problem Solving from Nature (PPSN'2). Brussels, Belgium, pp. 281–290, 1992.

[12] R. H. Storer, D. Wu, and R. Vaccari, "New search spaces for sequencing problems with application to job shop scheduling," Management Science, vol. 38, no. 10, pp. 1495–1509, 1992.

[13] E. D. Taillard, "Parallel taboo search techniques for the job shop scheduling problem," ORSA Journal on Computing, vol. 6, no. 2, pp. 108–117, 1994.

[14] E. Demirkol, S. Mehta, and R. Uzsoy, "A computational study of shifting bottleneck procedures for shop scheduling problems," Journal of Heuristics, vol. 3, no. 2, pp. 111–137, 1997.

[15] E. Balas and A. Vazacopoulos, "Guided Local Search with Shifting Bottleneck for Job Shop Scheduling," Management Science, vol. 44, no. 2, pp. 262–275, 1998.

[16] J. C. Beck, T. K. Feng, and J. Watson, "Combining constraint programming and local search for job-shop scheduling," INFORMS Journal on Computing, vol. 23, no. 1, pp. 1–14, 2011.

[17] J. P. Caldeira, "Private Communication of Result 2869 for ta62 to Éric D. Taillard, listed on Éric Taillard's Page," 2003.

[18] J. Carlier and E. Pinson, "An algorithm for solving the job-shop problem," Management Science, vol. 35, no. 2, pp. 164–176, 1989.

[19] J. F. Gonçalves and M. G. C. Resende, "An extended akers graphical method with a biased random-key genetic algorithm for job-shop scheduling," International Transactions on Operational Research, vol. 21, no. 2, pp. 215–246, 2014.

[20] A. Henning, " Praktische job-shop scheduling-probleme," Ph.D. thesis, Friedrich-Schiller-Universität Jena, Jena, Germany, 2002.

[21] E. Nowicki and C. Smutnicki, " A fast taboo search algorithm for the job shop problem," Management Science, vol. 42, no. 6, pp. 783–938, 1996.

[22] E. Nowicki and C. Smutnicki, " An advanced taboo search algorithm for the job shop problem," Journal of Scheduling, vol. 8, no. 2, pp. 145–159, 2005.

[23] P.M. Pardalos, O. V. Shylo, and A. Vazacopoulos, " Solving job shop scheduling problems utilizing the properties of backbone and big Valley," Computational Optimization and Applications, vol. 47, no. 1, pp. 61–76, 2010.

[24] B. Peng, Z. Lü, and T.C.E. Cheng, " A tabu search/path relinking algorithm to solve the job shop scheduling problem," Computers and Operations Research, Vol. 53, pp. 154–164, 2015.

[25] F. Pezzella and E. Merelli, " A tabu search method guided by shifting bottleneck for the job shop scheduling problem," European Journal of Operational Research, vol. 120, no. 2, pp. 297–310, 2000.

[26] O. V. Shylo and H. Shams, " Boosting Binary Optimization via Binary Classification: A Case Study of Job Shop Scheduling," cs.AI/math.OC abs/1808.10813, arXiv, 2018.

[27] P. Vilím, P. Laborie, and P. Shaw, " Failure-Directed Search for Constraint-Based Scheduling-Detailed Experimental Results," in CPAIOR'2015, Barcelona, Spain, pp. 437–453, 2015.

[28] T. Yamada and R. Nakano, " Genetic algorithms for job-shop scheduling problems," In Proceedings of Modern Heuristic for Decision Support, March18-19, 1997, London, England, UK, pp. 67–81, 1997.

[29] C. Zhang, Y. Rao, and P. Li, " An effective hybrid genetic algorithm for the job shop scheduling problem," International Journal of Advanced Manufacturing Technology, vol. 39, pp. 965–974, 2008.

[30] E. Bonabeau, M. Dorigo, and G. Theraulaz, " Swarm Intelligence: From Natural to Artificial Systems," Oxford University Press, New York, NY, USA, vol. 1, 1999.

[31] E. Koc, " The Bees Algorithm Theory, Improvements and Applications," Ph.D Thesis, Cardiff University, Cardiff, UK, 2010.

[32] D. Karaboga, "An Idea Based On Honey Bee Swarm for Numerical Optimization," Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, Vol. 200, 2005.

[33] V. Tereshko and A. Loengarov, "Collective decision making in honey-bee foraging dynamics," Computing and Information Systems, vol. 9, no. 3, pp. 1, 2005.

[34] V. Tereshko, "Reaction-diffusion model of a honeybee colony's foraging behaviour," In International Conference on Parallel Problem Solving from Nature, Springe, Berlin, Heidelberg, pp. 807–816, September 2000.

[35] V. Tereshko and T. Lee, "How information-mapping patterns determine foraging behaviour of a honey bee colony," Open Systems and Information Dynamics, vol. 9, no. 2, pp. 181–193, 2002.

[36] Y. Guo, X. Cao, H. Yin, and Z. Tang, "Coevolutionary optimization algorithm with dynamic sub-population size," International Journal of Innovative Computing, Information and Control, vol. 3, no. 2, pp. 435–448, 2007.

[37] M. Dorigo and T. Stützle, "Ant Colony Optimization," MIT Press, Cambridge, 2004. ISBN: 978-0-262-04219-2.

[38] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," IEEE Computational Intelligence Magazine, vol. 1, no. 4, pp. 28–39, 2006.

[39] V. Maniezzo and A. Carbonaro, "Ant colony optimization: An overview," In Essays and surveys in metaheuristics, Springer, US, pp. 469–492, 2002.

[40] T. Stützle, "Ant colony optimization," In International Conference on Evolutionary Multi-Criterion Optimization, Springer, Berlin, Heidelberg, pp. 2–2, April 2009.

[41] J. Kennedy, "Particle swarm optimization," In Encyclopedia of machine learning, Springer, US, pp. 760–766, 2011.

[42] S. C. Chu, P. W. Tsai, and J. S. Pan, "Cat swarm optimization," Proc. of the 9th Pacific Rim International Conference on Artificial Intelligence, LNAI 4099, pp. 854–858, 2006.

[43] M. Bakhouya and J. Gaber, "An Immune Inspired-based Optimization Algorithm: Application to the Traveling Salesman Problem," Advanced Modeling and Optimization, vol. 9, no. 1, pp. 105–116, 2007.

[44] K. M. Passino, "Biomimicry of Bacteria Foraging for Distributed Optimization and Control," IEEE Control Systems Magazine, vol. 22, pp. 52–67, 2002.

[45] K. N. Krishnanand and D. Ghose, "Glowworm swarm optimization for searching higher dimensional spaces," In Springer, Innovations in Swarm Intelligence, C. P. Lim, L. C. Jain and S. Dehuri, Eds. Heidelberg, 2009.

[46] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," Journal of Global Optimization, vol. 39, no. 3, pp. 459–471, 2007.

[47] D. Karaboga, B. Gorkemli, C. Ozturk and N. Karaboga, "A comprehensive survey: artificial bee colony (abc) algorithm and applications," Artif Intell Rev, vol. 42, no. 1, pp. 21–57, 2014.

[48] S. Ashrafinia, M. Naeem, and D. Lee, "Discrete Artificial Bee Colony for Computationally Efficient Symbol Detection in Multidevice STBC MIMO Systems," Advances in Artificial Intelligence, 2013.

[49] T.M. PanQ-K, P. Suganthan, and T. Chua, “A discrete artificial bee colony algorithm for the lot-streaming flow shop scheduling problem,” Inf Sci, vol. 181, pp. 2455–2468, 2011.

[50] A. Singh, “An artificial bee colony algorithm for the leaf constrained minimum spanning tree problem,” Appl Soft Comput, vol. 9, pp. 625–631, 2009.

[51] S. Sundar and A. Singh, “A swarm intelligence approach to the quadratic multiple knapsack problem,” In ICONIP 2010, Lecture notes in computer science, Springer, Berlin, vol 6443, pp. 626–633, 2010.

[52] T.M. Pan Q-K, P. Suganthan, and AH. L. Chen, “A discrete artificial bee colony algorithm for the total flowtime minimization in permutation flow shops,” Inf Sci, vol. 181, pp. 3459–3475, 2011.

[53] Z. Beheshti and S. M. Hj. Shamsuddin, “A Review of Population-based Meta-Heuristic Algorithms,” lnt. J. Advance. Soft Comput. Appl., vol. 5, no. 1, 2013.

[54] D. Karaboga and B. Akay, “A comparative study of artificial bee colony algorithm,” Appl Math Comput, vol. 214, pp. 108–32, 2009.

[55] D. T. Pham and M. Castellani, “The bees algorithm: Modelling foraging behaviour to solve continuous optimization problems,” Sage J., vol. 223, pp. 2919–2938, 2009.

[56] M. Dorigo and M. Birattari, “Swarm intelligence,” Scholarpedia, vol. 2, no. 9, pp. 1462, 2007.

[57] M. Dorigo, “Optimization, Learning and Natural Algorithms,”, PhD thesis, Dipartimento di Elettronica, Politecnico di Milano, Italy, pp. 140, 1992.

[58] M. Dorigo, V. Maniezzo, and A. Colorni, “The ant system: optimization by a colony of cooperating agents,” IEEE Transactions on Systems, Man, and Cybernetics-Part B , vol. 26, no 1, pp. 29–41, 1996.

[59] M. Dorigo, G. D. Caro, and L. M. Gambardella, “Ant algorithms for discrete optimization,” Artificial life, vol. 5, pp. 137–172, 1999.

[60] M. Tuba and R. Jovanovic, “An Analysis of Different Variations of Ant Colony Optimization to the Minimum Weight Vertex Cover Problem,” Transactions on Information Science and Applications, vol. 6, pp. 936–945, 2009.

[61] M. Grepinšek, S. H. Liu, and M. Mernik, ” Exploration and exploitation in evolutionary algorithms: A survey,” ACM Computing Surveys, vol. 45, no. 3, pp. 1-33, 2013.

[62] E. G. Talbi, “A taxonomy of hybrid metaheuristics, ” International Journal of Heuristics, vol. 8, no. 5, pp. 541–564, 2002.

[63] X. S. Yang, “Nature-Inspired Optimization Algorithms, ” Elsevier, London, 2014.

An IoT-based Fire Safety Management System for Educational Buildings: A Case Study

Souad Kamel[1]
Department of Computer and Network Engineering
College of Computer Sciences and Engineering
University of Jeddah, Jeddah 21959, Saudi Arabia

Amani Jamal[2]
Computer Science Department
Faculty of Computing and Information Technology
King Abdulaziz University, Jeddah 21589, Saudi Arabia

Kaouther Omri[3]
Department of Computer and Network Engineering
College of Computer Sciences and Engineering
University of Jeddah, Jeddah 21959, Saudi Arabia

Mashael Khayyat[4]
Department of Information Systems and Technology
College of Computer Sciences and Engineering
University of Jeddah, Jeddah 21959, Saudi Arabia

Abstract—**Safety is a serious concern that should be addressed carefully in different locations including homes, workplaces and educational buildings. The risk of fire is the most significant threat in many educational facilities such as schools, universities, offices, etc. The main goal of this work is developing an effective system that allows early managing of fires to avoid material and human losses. With the advent of the Internet of Things (IoT), the implementation of such system became possible. A low-cost system incorporating IoT sensors is constructed in this study to collect data (heat, the number of people at the fire scene, ...) in real time. The system provides a control panel that displays readings from all sensors on a single web page. When the collected values exceed a particular threshold, the system sends a message to the building keeper's phone, allowing him to notify the authorities or dispatch firemen in real time. One of the system's most important characteristics is that it keeps track of how many people are at the fire scene, simplifying the evacuation process and allowing civil defense authorities to efficiently manage resources. The system has been successfully tested in a variety of circumstances in an educational building (Al-Faisaliah female campus, University of Jeddah, Saudi Arabia).**

Keywords*—*Safety; fire; Internet of Things (IoT); sensors; cloud based platform; ThingSpeak

I. Introduction

Safety is defined as the condition of being protected from or unlikely to cause danger, risk, or injury. People are continuously exposed to dangers in their homes, workplace or roads. Occupants of Educational buildings such as students, educators and administrative staff have different awareness levels regarding safety practices and beliefs. Since educational buildings are continuously occupied, safety issues should be addressed properly and procedures used to reduce risks have to be implemented prior to unexpected accidents. The main objective of the present paper is to develop and implement a complete management system for monitoring safety. This can improve the safe conditions in educational buildings especially in Al-Faisaliah female campus, University of Jeddah, Saudi Arabia. To achieve the research objectives, a procedure including several steps will be implemented. The first step is dedicated to the study of safety requirements in educational buildings issued in official documents provided by civil defense authorities [1]. In the second step, to assess the commitment and awareness of the users of Al-Faisaliah Campus with health and safety conditions, an online survey was chosen to collect data because it can reach a large number of individuals and has a minimal risk of data inaccuracies.

Results of the survey conducted have shown that the safety culture inside Al-Faisaliah Campus is very poor. The results showed also that fire is the most dangerous threat that can occur. The official records have verified that fact. According to Saudi Civil Defense statistics [1], for the year 1440-1441 H, the number of firefighting operations was more than 42,000, equivalent to 119 firefighting operations per day. More than 14,000 were fires in the workplace, at a rate of 35.41%. The region of Mecca occupied the first place in the number of operations. Human losses are about 2000 cases, between 149 deaths and 1,809 injuries, and financial losses are more than 49 million Saudi Riyals (1 USD = 3.76 SAR).

The previous statistics have clearly indicated a rise in the daily rate of fires in workplaces such as schools and offices. The average rate of this kind of fires reached 42 fires per day. There are a multitude of causes for fires. According to the statistics, thermal demand was one of the most common causes of fires, with a percentage of 37.71%, equivalent to 45 fires per day. Next, 22.0% of the electricity is tampered with an average of 27 fires per day. Finally, the third cause of fire is the flaming heat source. From statistics above, the conclusion is that fires are big threats that may cause disasters in the studied educational building. In order to contribute to limit dangers, smart solutions relying on new technologies such as Internet of Things (IoT) should be developed and deployed. The proposed safety management system may help first educational building occupants in reacting immediately once a threat is initiated and second may assist civil defense authorities during their interventions. In addition, the safety department of the University should take benefit from the proposed system to remotely control the safety situation inside the campus.

Through using IoT technology, a complete low-cost smart solution that enables the management and monitoring of fire in real time was suggested. The building's status was remotely monitored via a friendly dashboard. This dashboard displays the data issued from sensors in one Web page. A message delivered to the building keeper's mobile phone is activated if a collected parameter exceeds its threshold. So that he can call the police or firefighters to arrive right away. A huge benefit of the system is its ability to count the number of individuals involved at the fire scene. This will ease the evacuation efforts.

The remaining of this paper is organized as follows. First, many studies related to safety issues including advantages and disadvantages of each approach will be summarized. Second, the proposed safety management system as well as its components and its operation will be detailed. finally, conclusions, recommendations and future work will be drawn.

II. Literature Review

A. Safety

This short literature review tries to cover the main aspects of safety that will be addressed in this study. Fire is considered as the most significant risk that has to be assessed regularly. In [2], qualitative approaches have been presented. The purpose was to identify and eliminate fire hazards. The study established that hotel utilities are of high risk. 76 items related to safety in hotels have been proposed and classified into seven main classes. The proposed methodical approach can be used by fire safety inspectors. Several tools including short oral survey, semi-structured interviews with subcontractors, etc. have been used to assess safety culture among investors working in the household construction sector in Australia [3]. Thus, a method of source causes analysis was employed to categorize the safety culture of subcontractors into seven different areas, including the building site, work procedures, equipment and materials. Recommendations drawn at the end of the study are built around free training of the subcontractors. The Cypriot manufacturing sector safety practices have been studied in [4] using a nation-wide survey in Nicosia. The opinion of managers regarding fire occurrence has been demonstrated to be influenced by several factors such as the no-smoking rule, the presence of fire alarms, and the practice of allowing employees to sleep on the job. Although the safety situation is generally good, there is still potential for enhancement, according to the study's findings. Since electricity is classified as a silent killer, identifying safety beliefs among Australian electrical workers has been addressed in [5]. The planned behavior concept has been used as a theoretical framework. Focus groups and interviews with 46 certified electrical professionals were analyzed according to advantages, disadvantages, referents, barriers and facilitators affecting respectively the safety beliefs and culture. In Saudi Arabia, few works have addressed the problem of safety. In reference [6], Saudi Arabia's safety regulations for worksites have been assessed by surveying the work being done on several projects. Protection assessment scores have been found to be typically superior across all categories for the large projects; whereas low assessment ratings, particularly in firefighting, healthcare, and comfort, were typical for small projects. Protection measures in residential building assessment procedures have been implemented . A field assessment study about fire awareness measures has been conducted in [7] through a survey in residential buildings. The obtained results showed that the safety awareness on fire is poor. Based on the observations, a number of strategies including effective codes and official requirements, designs taking into account safety and educational programs have been proposed as recommendation to improve safety awareness culture. In the same direction, the study in [8] has focused on the methods adopted by Saudi Arabia firms to solve safety issues when creating residential structures. The research concluded that safety design must be undertaken by qualified architects and engineers. Through a representative company survey conducted in Germany [9], workplace risk assessments have been studied. The frequency of patterns influencing OHS measures has been evaluated in N=6500 companies. The study suggested that more effort is to be deployed by the authorities to improve the safety practices and beliefs. In [10], it has been evaluated how the Hail region of Saudi Arabia views and employs electrical safety. Hail region level of electric safety awareness has been found to be 0.76 out of 4. This low score reflects a bad culture of electrical safety. Numerous suggestions covering numerous relevant parties have been put forth. From the above literature review about safety, it can be remarked that fire is the main concern since it may cause irreversible dangers that can affect human lives and properties in addition to its high impact on economy and social life [11]. For this reason, the study main focus will be fire safety management based on IoT technology.

B. Internet of Things (IoT)

The term "Internet of Things" (IoT) refers to a network of physical objects which are equipped with sensors, programs, as well as other tools that allow them to communicate with other objects and systems over the Internet. IoT is essential for raising living standards. Indeed, it is able to human well-being and the quality of life enhancement [12]. Fire is among the common problems that may be managed/solved by using solutions around IoT technology. In order to prevent the loss of priceless lives and critical infrastructure in the case of a fire, it is imperative to establish an early, and precise fire detection system. Integrating contemporary technology, such as IoT, advanced analytics, and WSN, can result in accurate fire detection systems for real-time monitoring and crisis management [13].

The abundance of sensors, which are small devices with environmental sensing capabilities, is what gives IoT its greatest strength. Technologies for detecting fire can be divided into those that detect temperature, gases, and flames. Early-fire detection is the main function of fire sensing devices. A good fire system imposes that sensors must Identify a fire issue in its early stages. Heat sensors function properly. But, they are not fast. Make them moving can increase their speed. Smoke detectors have a poor accuracy rate. They can perform better if a visual sensor system is added. Due to the irreversible nature, fragility, and poor selectivity of gas sensors, their usage is extremely restricted in buildings.

Currently, the emphasis is on employing robots to fight fire in critical cases. This procedure is always performed from the outside of the burning place. But robots are heavier due to sensor systems and fire suppression equipment mounted on them. This fact creates a difficulty with balance and high-speed movement for internal fire detection. Hence, more study is required to develop improved sensing systems [14].

C. Related Work

An review of various relevant Internet - of - things intelligent fire detection and management works is provided in this section. Reference [15] utilized numerous sensors to gather real-time readings. When an emergency arises, sensor data are examined to start a sprinkler system. This has an important issue which is the non-prevention of fire occurrence. The authors in [16] used a Raspberry Pi to create a fully coded fire warning framework. Whenever a fire is discovered, a tailored app sends out an alert along with a URL to a website that contains pictures captured by embedded cameras. A Convolutional Neural Network approach for identifying fire in real pictures was introduced in [17]. Results were superior than those suggested in the literature. The authors proposed a future improvement that would use videos rather than photos. The authors in [18] proposed a system that uses multiple sensors to collect readings. An artificial intelligence -based algorithm analyzes and processes the gathered data. In the event of fires, airflow and a water spray are then activated. This has a significant flaw since it can only function properly in enclosed spaces. The development of a smart ventilation and lighting solution that could recognize people and control lights was the subject of reference [19]. Additionally, heat and gases could be remotely measured and gathered. The fact that this equipment was only used in one chamber is among the drawbacks. More improvements and research may be required before it can be used on a wide basis. The Ubidots platform was utilized by the authors of [20] to create an improved forest fire monitoring system. A buzzer sounds to notify users whenever sensor readings cross a predetermined limit. One problem with the suggested solution is that sensors placed outside are inaccurate since weather conditions can alter them. Authors created a smart home inspection for fire prevention in [21] that incorporated a multitude of sensors in each connected home area. To quickly alert the user of a fire incident, they used the Global System for Mobile Communications (GSM). Experiment is performed using a Fire Dynamic Simulator. One problem with the suggested solution is how to effectively handle the enormous volume of data that has been gathered. Some of the proposed systems above include technical flaws, such as needless complexity and exorbitant costs. Therefore, this research suggests an intelligent system that provides real-time data collection and monitoring and alerting for the building occupants and any concerned authorities.

In order to make the system innovative and better than the previously developed systems, a comparative study based on several criteria has been conducted (Table I). The main criteria considered in this study are the types of:

- sensors
- the environment
- alarms
- dashboard
- new features
- processing (ML/IP/FL: ML for machine learning. IP for image processing. And IF for fuzzy logic)

Based on the previous comparison (Table I), the following conclusions can be noted:

TABLE I. Comparison between the Proposed System and Similar Ones

Criteria / Reference	[15]	[16]	[17]	[18]	[19]	[20]	[21]	The proposed system
Temperature	√	×	×	√	√	√	√	√
Smoke	√	×	×	√	√	×	√	√
Flame	×	√	×	√	√	×	√	×
Gas	√	×	×	√	√	×	√	×
PIR	×	×	×	×	×	√	×	√
Camera	×	√	√	×	×	×	×	×
Cloud dash-board	√	×	×	×	×	√	×	√
Indoor	√	√	√	√	√	×	×	√
ML/IP/FL	×	√	√	√	×	×	√	×
Counting People	×	×	×	×	×	×	×	√
alert	×	√	×	√	√	√	√	√

- Many of the systems above have technical limitations (unnecessary complexity).
- The largest error is trusting a single type of fire sensor. Doing so will lead to more false alerts.
- It is insufficient to just use an audible alarm to notify stakeholders. It may be necessary to warn the building guard via message whenever a fire arises in particular circumstances, such as when he is outdoors.
- Sending a message is not enough in case of emergencies. A dashboard is required to the surveillance of the building.
- Firefighters frequently don't know how many people are in the burning building. This makes the evacuation process exceedingly challenging. By keeping track of this number, fireman's work will be easily achieved and their resources will be used more effectively.

In this prototype, a multitude of sensors were used. This can ensure precise fire detection. Additionally, a buzzer and LED were employed to create an audible and a visual alarms in case of emergency. Then, two Passive Infrared (PIR) sensors for counting people entering and exiting were deployed. Finally, a dashboard that enables the surveillance of the building was implemented. When an emergency arises, a message will be automatically delivered to the building guard.

III. Design and Development

A. Study Area

The study area chosen for assessing performances of this low-cost fire monitoring station is Al-Faisaliah Campus (Female branch of the University of Jeddah: Fig. 1). More precisely, the experiments are conducted in Building 11 (see

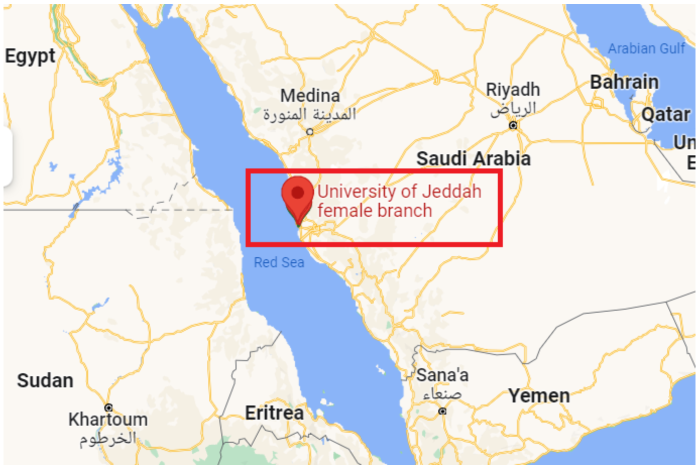


Fig. 1. Al-Faisaliah Campus (Female Branch of the University of Jeddah).

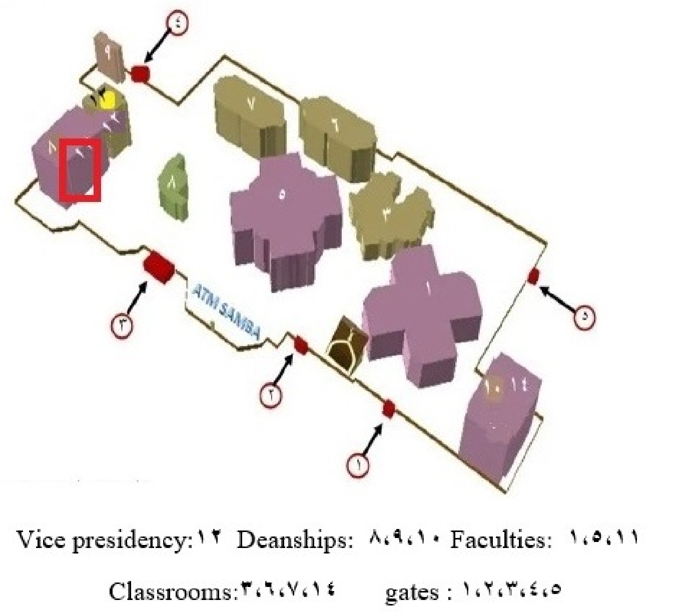

ATM SAMBA
Vice presidency:١٢ Deanships: ٨،٩،١٠ Faculties: ١،٥،١١
Classrooms:٣،٦،٧،١٤ gates : ١،٢،٣،٤،٥

Fig. 2. The Monitored Area (Building 11: in Red Square).

Fig. 2). For the development of this smart fire monitoring system, the work was divided into the following stages:

B. Functional and Non-Functional Requirements

The functional requirements specify what the proposed solution must perform during its operation. On the other hand, non-functional requirements are those characteristics of the device that can be observed throughout its execution. The following are the specified requirements:

1) Functional Requirements: The stakeholders are the building guard and the administrator.

- the building guard can view:
  1) sensors readings
  2) the number of persons in the fireplace
  3) any sudden changes in the monitored zone
- The administrator can determine:
  1) the temperature at which an alarm should be triggered
  2) the humidity which an alarm should be triggered
  3) the smoke t which an alarm should be triggered
- The system will be able to:
  1) display the dashboard
  2) determine how many people are in a building fire.
  3) issue different alerts during the early stages of a fire emergency

2) Non-Functional Requirements:

- Usability: The system shall be easy and simple to learn and use
- Portability: The system can be used in any indoor environment such as buildings, workplaces, hospitals, etc.
- Accuracy: The system should be accurate to avoid any faulty fire alarm.

C. Hardware Requirements

To make a prototype of an internet-of-things enabled intelligent fire surveillance system, one needs a microcontroller [22], sensors and a wifi module [23] for sending data from sensors to the internet. In this case, a micro controller integrating wifi which is NodeMCU [24] was chosen. Also, sensors [temperature (DHT11) [25], smoke (MQ2) [26]], a light emitting diode (LED) [27], a buzzer [28] and jumper wires [29] were deployed.

D. Block Diagram

The block diagram below serves as the foundation for the intelligent fire continuous monitoring system (Fig. 3).

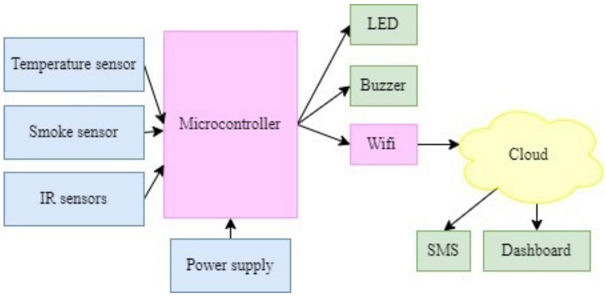


Fig. 3. Block Diagram.

The data issued from temperature, IR and gas sensors deployed in the monitored area are collected. Then transmitted to the microcontroller. Through wifi, sensors readings are displayed on a user-friendly dashboard. When readings are high and will cause a fire (cross a threshold), an alert message is sent to the building guard, a buzzer will beep accordingly and the LED will glow.

E. Circuit

Fig. 4 shows the circuit of the proposed smart system.

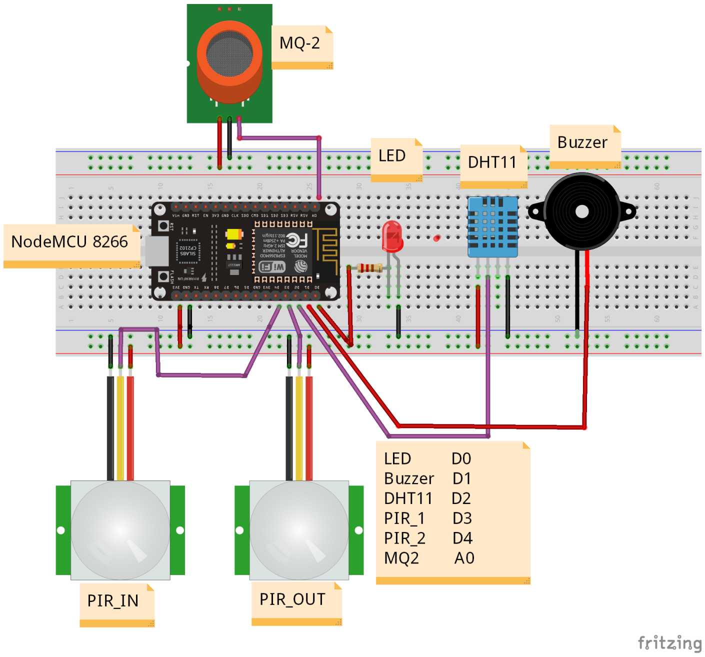


Fig. 4. Circuit Diagram.

F. Prototype

Fig. 5, Fig. 6, and Fig. 7 are a depiction of the exterior view, the exterior view without roof, and the interior view of the finalized system prototype, respectively.

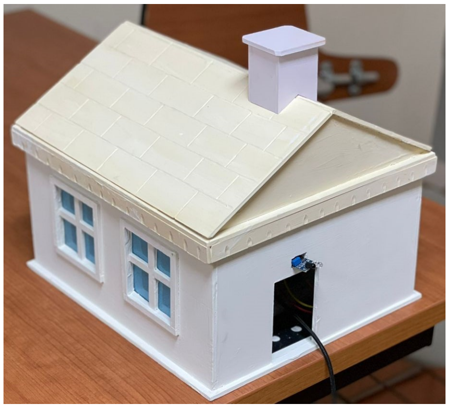


Fig. 5. View of the Developed Prototype from the Outside.

G. Fire Detection

According to Algorithm 1, temperature and smoke data are collected, sent to the cloud. End users are notified in the event that any parameter has abnormally high level.

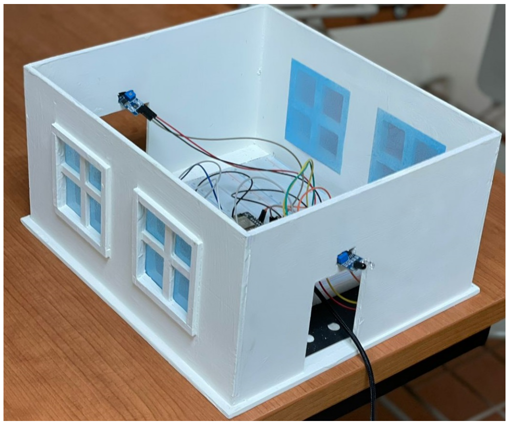


Fig. 6. View of the Developed Prototype from the Outside (without Roof).

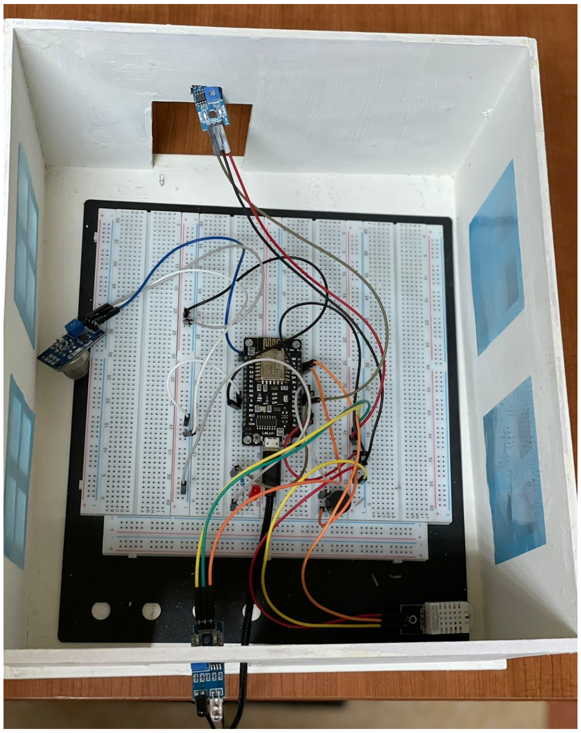


Fig. 7. A View of the Developed Prototype's Interior

IV. Results and Discussion

To visualize sensors' readings, Thingspeak [30] was used. ThingSpeak is a cloud service. This IoT analytics platform solution is really effective. The examination of real-time data streams is possible. Devices can continuously submit data to ThingSpeak. This can make instantaneous visualization of live data. It can also send alerts in emergency cases. The graphs below (Fig. 8, Fig. 9) show the results from the measured values taken from used sensors during an experiment conducted on site. In case of fire, an SMS is sent to the building guard automatically which can let him contacting firefighters immediately (Fig. 10).

Algorithm 1 Fire Detection

```
Require: Temperature sensor, Smoke sensor
Ensure: Warning notification
  Initialization:
  Set threshold for temperature $T_{TH}$
  Set threshold for smoke $S_{TH}$
  Capture temperature measured $T$ from target environment
  Capture smoke measured $S$ from target environment
  if ($T >= T_{TH}$) AND ($S >= S_{TH}$) then
    Red light (LED)
    sound alarm (Buzzer)
    notify the building guard (SMS)
  end if
```

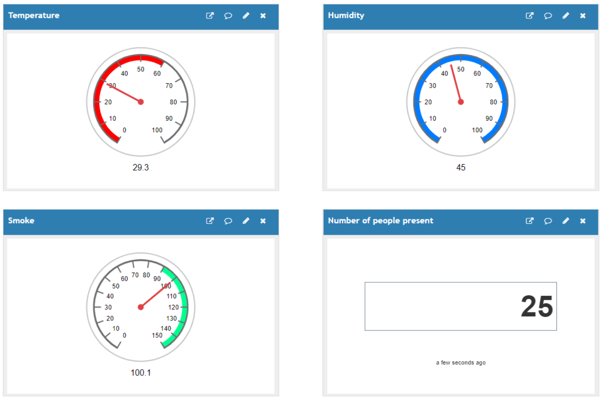


Fig. 8. Screenshot of the Continuous Monitoring in ThingSpeak.

V. Conclusion and Future Work

A cost effective smart fire system that provides real-time monitoring and data collection was deployed in this work. The suggested solution enables remote building status monitoring via a user-friendly dashboard that compiles sensor data into a single web page. A buzzer, a red light, and a message sent to the building keeper's mobile phone inform him whenever a collected parameter exceeds its threshold so that he can call authorities or firemen for assistance right away. The proposed technique has the important benefit of counting the number of persons at the building fire, which will speed up the evacuation procedure. Different scenarios have been successfully tested with the suggested system. The addition of an image processing method with a camera and real-time data analysis from sensors utilizing sophisticated algorithms can be considered as a future project. Another future work can be built around the deployment of a wireless sensor network to cover all the buildings existing in the campus. This can be performed by taking the advantages of the tremendous development that wireless communication technologies are experiencing today.

Acknowledgment

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-02-077-DR). The authors, therefore, acknowledge with thanks the University of Jeddah technical and financial support.

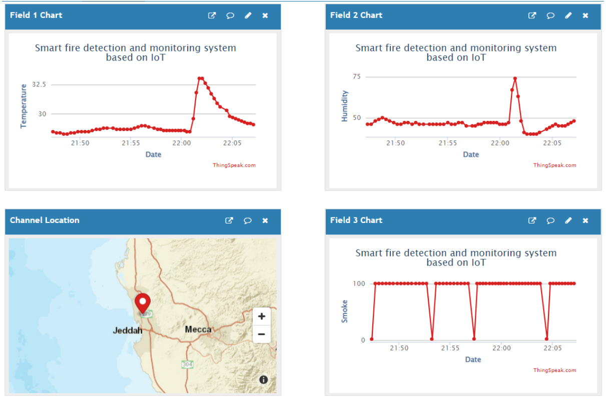
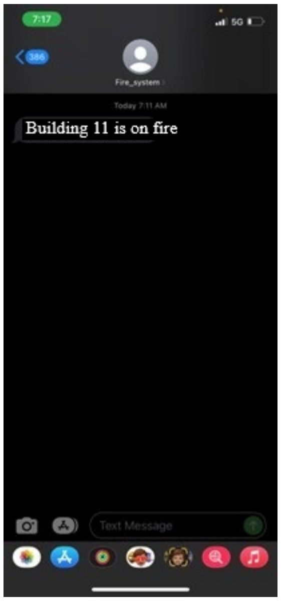


Fig. 9. Screenshot of Graphs Showing Sensors Readings.



Today 7:11 AM

Building 11 is on fire

Fig. 10. SMS Sent in Case of Fire.

References

[1] C. defence-report (1440-1441), "Civil defence report (1440-1441)." [Online]. Available: https://998.gov.sa/Ar/Marquee/pages/statistics.aspx

[2] M. A. Hassanain, "Approaches to qualitative fire safety risk assessment in hotel facilities," *Structural Survey*, 2009.

[3] P. Wadick, "Safety culture among subcontractors in the domestic housing construction industry," *Structural Survey*, 2010.

[4] G. Boustras, R. Bratskas, V. Tokakis, and A. Efstathiades, "Safety awareness of practitioners in the cypriot manufacturing sector," *Journal of Engineering, Design and Technology*, 2011.

[5] K. M. White, N. L. Jimmieson, P. L. Obst, P. Gee, L. Haneman, B. O'Brien-McInally, and W. Cockshaw, "Identifying safety beliefs among australian electrical workers," *Safety science*, vol. 82, pp. 164–173, 2016.

[6] M. O. Jannadi and S. Assaf, "Safety assessment in the built environment of saudi arabia," *Safety Science*, vol. 29, no. 1, pp. 15–24, 1998.

[7] M. S. Al-Homoud, A. A. Abdou, and M. M. Khan, "Safety design practices in residential buildings in saudi arabia," *Building Research & Information*, vol. 32, no. 6, pp. 538–543, 2004.

[8] M. S. Al-Homoud and M. M. Khan, "Assessing safety measures in residential buildings in saudi arabia," *Building Research & Information*, vol. 32, no. 4, pp. 300–305, 2004.

[9] U. Lenhardt and D. Beck, "Prevalence and quality of workplace risk assessments–findings from a representative company survey in germany," *Safety Science*, vol. 86, pp. 48–56, 2016.

[10] S. Boubaker, S. Mekni, and H. Jerbi, "Assessment of electrical safety beliefs and practices: A case study," *Engineering, Technology & Applied Science Research*, vol. 7, no. 6, pp. 2231–2235, 2017.

[11] N. Nadzim and M. Taib, "Appraisal of fire safety management systems at educational buildings," in *SHS Web of Conferences*, vol. 11. EDP Sciences, 2014, p. 01005.

[12] S. Smys, "A survey on internet of things (iot) based smart systems," *Journal of ISMAC*, vol. 2, no. 04, pp. 181–189, 2020.

[13] A. Mukherjee, S. K. Shome, and P. Bhattacharjee, "Survey on internet of things based intelligent wireless sensor network for fire detection system in building," in *Communication and Control for Robotic Systems*. Springer, 2022, pp. 193–200.

[14] A. Gaur, A. Singh, A. Kumar, K. S. Kulkarni, S. Lala, K. Kapoor, V. Srivastava, A. Kumar, and S. C. Mukhopadhyay, "Fire sensing technologies: A review," *IEEE Sensors Journal*, vol. 19, no. 9, pp. 3191–3202, 2019.

[15] H. Alqourabah, A. Muneer, and S. M. Fati, "A smart fire detection system using iot technology with automatic water sprinkler." *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 11, no. 4, 2021.

[16] K. R. Ahmed and M. A. Hossain, "An automated iot based fire detection & safety control for garments industry in bangladesh: A-study," in *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*. IEEE, 2021, pp. 1–5.

[17] S. Majid, F. Alenezi, S. Masood, M. Ahmad, E. S. Gündüz, and K. Polat, "Attention based cnn model for fire detection and localization in real-world images," *Expert Systems with Applications*, vol. 189, p. 116114, 2022.

[18] A. Rehman, M. A. Qureshi, T. Ali, M. Irfan, S. Abdullah, S. Yasin, U. Draz, A. Glowacz, G. Nowakowski, A. Alghamdi *et al.*, "Smart fire detection and deterrent system for human savior by using internet of things (iot)," *Energies*, vol. 14, no. 17, p. 5500, 2021.

[19] M. Mahbub, M. M. Hossain, and M. S. A. Gazi, "Cloud-enabled iot-based embedded system and software for intelligent indoor lighting, ventilation, early stage fire detection and prevention," *Computer Networks*, vol. 184, p. 107673, 2021.

[20] P. Kanakaraja, P. S. Sundar, N. Vaishnavi, S. G. K. Reddy, and G. S. Manikanta, "Iot enabled advanced forest fire detecting and monitoring on ubidots platform," *Materials Today: Proceedings*, vol. 46, pp. 3907–3914, 2021.

[21] F. Saeed, A. Paul, A. Rehman, W. H. Hong, and H. Seo, "Iot-based intelligent modeling of smart home environment for fire prevention and safety," *Journal of Sensor and Actuator Networks*, vol. 7, no. 1, p. 11, 2018.

[22] S. F. Barrett, "Arduino microcontroller processing for everyone!" *Synthesis Lectures on Digital Circuits and Systems*, vol. 8, no. 4, pp. 1–513, 2013.

[23] C. Bell, "Introducing the arduino," in *Beginning IoT Projects*. Springer, 2021, pp. 31–70.

[24] J. Desai, P. Moga, and R. Patwardhan, "Design and implementation of iot based smart library," *International Research Journal of Innovations in Engineering and Technology*, vol. 6, no. 1, p. 19, 2022.

[25] W. Gay, "Dht11 sensor," in *Advanced Raspberry Pi*. Springer, 2018, pp. 399–418.

[26] S. Shrestha, V. K. Anne, and R. Chaitanya, "Iot based smart gas management system," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2019, pp. 550–555.

[27] G. Held, *Introduction to light emitting diode technology and applications*. Auerbach publications, 2016.

[28] J. Christian, N. Komar *et al.*, "Prototipe sistem pendeteksi kebocoran gas lpg menggunakan sensor gas mq2, board arduino duemilanove, buzzer, dan arduino gsm shield pada pt. alfa retailindo (carrefour pasar minggu)," *Jurnal TICom*, vol. 2, no. 1, p. 92830, 2013.

[29] B. Stewart, *Adventures in Arduino*. John Wiley & Sons, 2015.

[30] C. Bell, "Using thingspeak," in *Beginning IoT Projects*. Springer, 2021, pp. 777–845.

Structural Vetting of Academic Proposals

Opeoluwa Iwashokun[1]
Department of Applied Information Systems
College of Business and Economics, University of Johannesburg
Johannesburg, South Africa

Abejide Ade-Ibijola[2]
Research Group on Data, Artificial
Intelligence, and Innovations for Digital Transformation
Johannesburg Business School, University of Johannesburg
Johannesburg, South Africa

***Abstract*—Increasing postgraduate enrollments gives rise to many proposal documents required for vetting and human supervision. Reading and comprehension of large documents is a boring and somewhat difficult task for humans which can be delegated to machines. One way of assisting supervisors with this routine screening of academic proposals is to provide an artificial intelligent (AI) tool for initial *structural vetting* — checking if sections of proposals are complete and appear where they are supposed to. Natural Language Processing (NLP) techniques in AI for document vetting has been applied in legal and financial domains. However, in academia, available tools only perform tasks such as checking proposals for plagiarism, spellings or grammar, word editing, and not structural vetting of academic proposal. This paper presents a tool named Auto-proofreader that attempts to perform the task of structural document review of proposals on behalf of the human expert using formal techniques and document structure understanding hinged on context free grammar rules (CFGs). The experimental results on a corpus of 20 academic proposals using confusion matrix technique for evaluation gives an overall of 87% accuracy. This tool is expected to be a useful aid in postgraduate supervision for vetting students' academic proposals.**

***Keywords*—Document structure; context free grammar; postgraduate supervision; artificial intelligence; natural language processing**

I. Introduction

Natural language processing (NLP) has been applied in document vetting across domains such as legal practice [1], [2], [3]. These areas of research in extracting information, text summarising and text vetting of documents is a difficult task for humans when several pages of a document or many documents are involved or short time is available. Daramola [4] observed that postgraduate supervisors are faced with the salient task of vetting many students' proposals to conform to certain academic standards, amidst other key roles in the University and must find the right balance for work and effectiveness. In recent times, South Africa experienced an increasing headcount in the number of post-graduate students' enrollment, impacting on the process of screening of submitted proposals [5]. There are yet many sub-standard proposals submissions by postgraduate student novice writers which is impacting negatively on the screening, feedback and vetting time of the assigned supervisors [4], [6]. The timeliness and effectiveness of screening these proposals can be assisted by an intelligent tool performing the specific task of vetting proposals based on prescribed proposal format guideline constructed as context free grammar (CFG) rules.

A. Challenges Facing Supervisors in the Process of Vetting Academic Proposals

Supervision is a critical component of postgraduate studies and many supervisors continue to grapple with promoting research ideas in students' academic writing and students' writing standards [7]; especially when they are still novice writers and in the first year of research writing [4]. There is an increasing pressure on postgraduate supervision as the number of enrollments continue to increase exponentially and universities are under more pressure for more quality research output [8]. Supervisors continue to assess the preparedness and candidature of student enrollees with the vetting outcomes of their proposals. They observed that students often submit proposals that are unacceptable (often too long or too short, poorly written or not well organised and often *missing critical proposal sections*). This is attributed to students not reading or understanding the guideline/instruction format or other peculiar reasons [4], [7]. Supervisors are expected to play a multi-faceted role in their discharge of duties. They are stretched thin as they support students' handling responsibilities and other academic responsibilities for the university [4], [9]. A better approach is to reduce drudgery by introducing technological tools for replacing traditional approach in supervising students [10] and specifically for vetting the structure of academic proposals.

B. What has been done?

1) Support for Students to improve Writings before Proposal Submission: There are lots of approaches and support programmes for scholars to develop writing techniques in the best way possible. There are provisions for writing workshops [11], writing groups [11], informal and online support services [11], mentoring programmes [12], writing editors, various grammar and spell-check productivity tools [13], [14]. These were supports for scholars to improve their writings which is as important as support for supervision. The findings of a review of students and supervisors by Hey-Cunningham *et al.* [10] explained that providing innovative solutions to improve the feedback mechanism in supervision is very important.

2) Support Toolkit for Supervisors: A survey conducted by Hey-Cunningham *et al.* [10] expresses that the theme common to many supervisors was the need for enhancements of supervisory approaches to academic writing standards for which the authors proposed innovations for a timely and effective feedback in supervision. Supervisors engage many generalised tools (e.g. plagiarism checker, editor review tools and grammar cum spell check error tools) to provide revision

TABLE I. CATEGORY OF EXISTING TOOLS FOR POSTGRADUATE SUPERVISION

Supervision tool	**Functions**
Plagiarism checkers e.g. Turnitin	It performs only text extraction for similarity check index against other literature. It is not a self-check document tool.
Feedback review tools e.g. Microsoft's word reviewer	It can assist a human reviewer to perform a document self-check, but it can be time-consuming especially when reviewing proposals of many students.
Grammar and spell check tools e.g. Grammarly, document proWriting aid, etc.	It can only assist the human reviewer to proofread the content of a proposal but not the structure or format layout of the proposal document the proposal document.

supports when screening submitted proposals of students. The perceived functions of these tools are laid out in Table I

C. The New Norm of Technological Aids in Postgraduate Supervision

Productivity tools are replacing and improving traditional and unconventional methods of performing higher education supervision [15], [16]. The effect of COVID-19 pandemic has even made it more necessary. For instance, one-to-one supervision meetings now commonly take place virtually and proposal document review process are more commonly done with various online collaborative productivity tools [17], [18], [16]. Productivity tools has been very effective in engaging dialogue between supervisors and their students, but has been proven to be time-consuming and tiresome when examining a large batch of students' proposals on a computer [19]. Many grammar and spell-check tools are also used by students and supervisors for fine-tuning grammar and editing spelling errors that may not be easily tracked by the eyes of a human reviewer. Popular examples are Grammarly, Microsoft word spell check, pro-Writing aid and language tool.

D. Gap

Researchers have used various forms of text processing technique to automatically extract and analyse documents such as business documents [20], clinical notes [21], legal documents [2], [22] and so on. NLP techniques have been used to perform text information extraction [23], named entity recognition [24], language to SQL translator [25], [26], summarisation [27], classification and examination of other textual contents such as CVs [28], invoices [20] and social media texts [29]. These NLP techniques and others have been largely used around the text content of a document, and sometimes short-text based documents. We considered document understanding an AI task and we have found no tool for vetting of large-content academic text based document such as proposals or similar academic writings. The question then is: *"how can we aid the vetting of academic proposals using existing NLP techniques?"*

E. Proposed Solution

In this paper, we have designed an approach which breaks up a proposal document into `tokens` that are basic recognisable `symbolic parts` of an academic proposal document. We also determine if the input proposal document parts satisfies a valid structure, defined in a proposal guide, by constructing a CFG for the acceptance of a valid proposal structure and REGEX to accept valid terminal symbols. A simple parse tree representation for a proposal document is given in Fig. 1. The document itself as the root node of the parse tree contains Section parts of a proposal document. The algorithm for document parsing is implemented using an existing PDF library named `iTextSharp` Java PDF library.

Similar technique was used for CV slicing [28], metadata extraction of PDF books [30], meta-analysis of clinical notes [21], business invoice document processing [20] and legal documents [27].

F. Contribution to Knowledge

This paper contributes to knowledge in the following ways:

1) the production of a simplistic CFG for recognising an academic proposal structure,
2) design of an approach for the automatic discovery of the document's structure of academic proposals, useful for other large text based document,
3) it promotes further research on automatic slicing of text documents using grammar based rules,
4) design of an algorithm for end-to-end automatic vetting of the logical sections and elements (i.e. structure) of an academic proposal and
5) it describes the implementation and evaluation of a software tool for vetting of academic proposals.

The rest of this document is organised as follows. Section II explains some background to this research problem and related research efforts in text documents comprehension. Section III contains the design concept for this work, while Section IV shows the result of the test of the software tool on proposal documents and the output results generated. An evaluation of these results is done in Section V, while we conclude and state further future work in Section VI.

II. BACKGROUND AND RELATED WORK

There is no directly related research work in vetting academic proposals. This section only explains various related NLP techniques and examples in past research for information extraction and summarisation of documents in general. Then highlights on the formal grammar approaches that has been applied by various researchers for automatic document discovery of large text documents. bluelightAnd lastly discusses some existing proofreading tools.

A. Automatic Document Discovery

NLP has gained a lot of popularity using computational methods to process spoken or written form of text by humans, with applied use cases in various business and enterprise based applications [31]. This field in artificial intelligence (AI) is gaining a lot of momentum and one key applied use is in the area of text processing for information extraction, summarisation or classification [23]. It is used in legal field for similar case matching and text summarisation of legal documents [2]. NLP techniques has been applied successfully for automatic comprehension of clinical notes [21], slicing and information extraction from curriculum vitae (i.e. CVs) [28], understanding

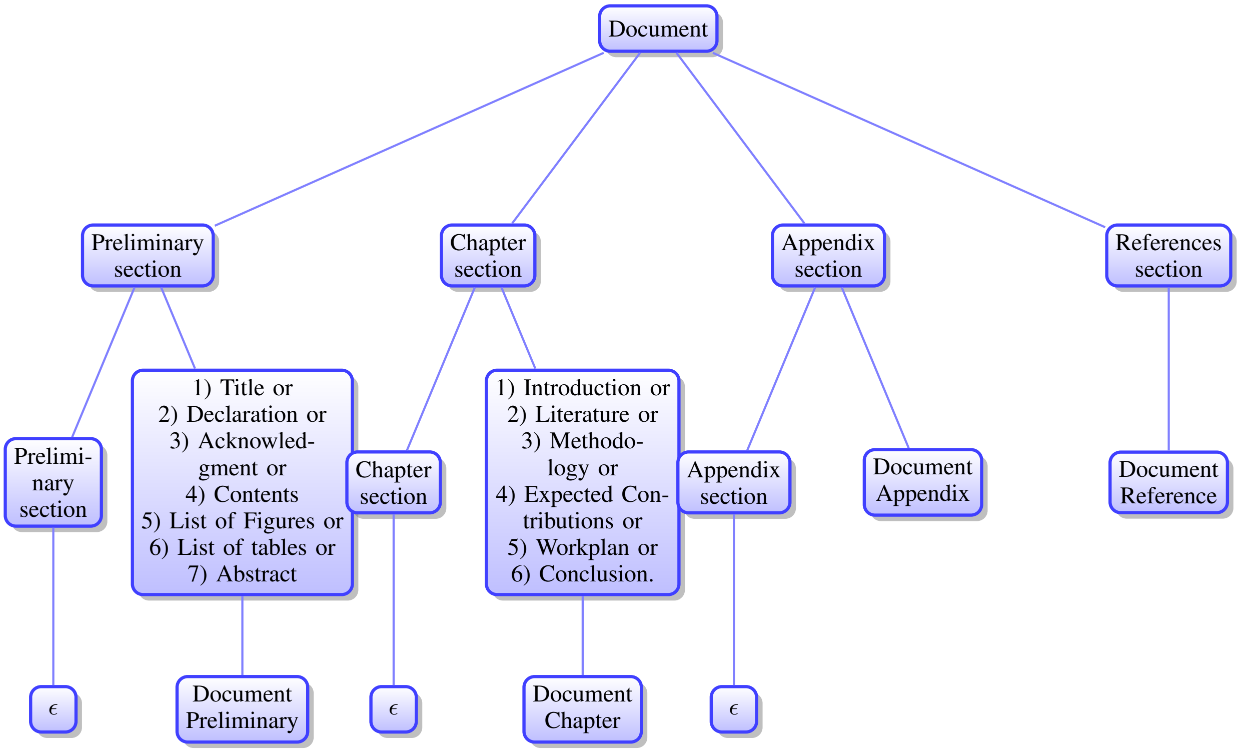


Fig. 1. Proposal Document Parse Tree.

and improvement of requirements documents [32], extracting parts of business (invoice) documents [20], finance chat messaging [33] and so on.

B. Related Work

1) Document Vetting of Legal Proceedings Document: Legal expert systems that use NLP for relevant information extraction and summarization for legal consumption now exist. The computational language processing of legal text documents has been very useful in drafting and analysing legal documents, classifying documents based on relevance to legal case and legal documents discovery and legal citations extractions [2], [3], [27]. More recently, a process tagged "Technology Assisted Review" is associated with the legal profession focused on categorizing legal documents and files based on relevance to case or legal information required [3]. It has become popular and replaced manual review of documents in the legal profession with a more effective automated approach.

2) Automatic Extraction from Business Documents: Glenda and Shilpa [1], implemented an NLP based document vetting process for Banks thereby reducing staff work load and increasing efficiency. NLP techniques have been successfully applied to processing business invoice documents [20], IT Projects system requirements document vetting [32] and automatic comprehension of business finance chats [33]. The applied use of NLP for document processing in finance has become very rapid and important.

3) Automatic Comprehension of Clinical Notes: Modern medicine has embraced NLP techniques for systematic reviews of several clinical trials with great measures of success, known as NLP-enhanced clinical trials research [34]. In a similar vein, a simple NLP translator was designed to decrypt clinical notes, creating friendly user plain texts from complex medical reports [21].

4) Automatic Slicing of CVs: Curriculum Vitae (CV) and resumes are structured documents that contain certain key elements information such as work experiences that talent-hunt specialist usually look out for in CVs during a job advert placement. Emil St. *et al.* [35] applied NLP text extraction techniques on several CV documents to determine candidate professional qualifications, which is useful for review and ease of vetting the relevance of the CV to the role advertised . At another instance, an NLP-based tool was designed to extract the logical sections of CVs using a set of CFG rules [28].

C. Some Background on Document Structure Parsing

The order of arrangement of a document largely explains its structure and can be described using a tree model or an hierarchical pattern [36], [37]. Anjewierden's [36] approach extracted the characteristics of text fonts and lines of text to discover the structure or document style by clustering tokens (in this case, a non-space characters or strings of the English alphabet identified in PDF document) into document elements identifiable on a document page. This was achieved using the characteristics of the token's dimensional co-ordinates on the document, which are upper left position co-ordinates of token and bottom right position co-ordinates on a document's page. These characteristics are syntactically analysed using rules by categorizing text into chunks of meaningful elements of the document that reveals the document structure. According to Anjewierden [36], a set of document object texts is sorted on

TABLE II. Existing Proofreading Tools and Techniques

	Tool	Main function	Technique
1.	Grammarly [38], [39], [40], [41]	Grammar, style and spell check	It combines rules, patterns and AI techniques in machine learning and deep learning
2.	Pro-write Aid [42]	Grammar, style and spell check	AI techniques in machine learning techniques
3.	Typely	Grammar, style and spell check	AI techniques in machine learning techniques aimed at high precision
4.	Custom-built proofreader tool e.g Chinese text automatic proofreader [43]	Grammar, style and spell check	entity recognition and Knowledge graph

their co-ordinate positions on a document page, then parsed through a set of shallow grammars to detect specific elements of the document's logical structure.

D. Discussion of Some Existing Proofreading Tools

There are quite a number of existing proofreader tools available as add-on or online tools for general writing problems on grammars, spellings and styles. Grammarly is quite popular for automatic proofreading in academic writing [38], [39], [40], [41]. Karyuatry and Rizqan [40] explains that Grammarly provides feedback on grammar errors and styling mistakes based on similarity patterns in real time. It is mainly available as a web based tool and built on AI system of a variety of natural language processing (NLP) statistic and machine learning based techniques. Pro-writing Aid is a similar proofreading tool for corrections in punctuation, grammar and style [42]. It is also available online and provides teachers and or students with feedback reports, with which they can improve writing. Proofreaders generally rely on AI techniques to come up with suggested corrections as feedback. Table II summarizes the functions and techniques of some existing tools. The table shows that existing proofreader tools have created a niche for improving writing by providing the feedback on grammar, punctuation and style. However, the novelty of our proposed tool seeks to provide feedback based on the organization (i.e. structure) and sections of an academic research proposal.

E. Definition of Terms

Definition 1: The document logical structure [36] is the layout for its constituents consisting of paragraphs, item lists, sections, tables etc which is easily identified by humans but has to be discovered computationally.

Definition 2: The document (logical) elements [36] is the proper thematic units that can be annotated and form part of the document logical structure, it is also referred to as document text segment. We can identify headings, sub-headings, paragraphs, tables, lists of information, page numbers and headers or footers as structural elements of a document.

Definition 3: A context free grammar (CFG) [44], [45] is a given grammar G defined by a four tuple given as
$G = (N, \Sigma, P, S)$ where:

1) N is a set of non-terminals,
2) Σ is a finite set of terminal symbols, which are the nodes of the grammar (such as symbols, alphabets and numbers),
3) P is a set of Productions and
4) S is a non-terminal start symbol.

.

Definition 4: Regular languages (RL) and regular expressions (REGEX) [44], [45] denotes regular languages. Both are represented by formulas involving the operations of concatenation, union and Kleene star. Regular languages are such languages that can be accepted by a finite automation. In formal terms, a regular language (L) for a given grammar over an alphabet set is Σ defined as any of the following:

1) a singleton language a where $a \in \Sigma$,
2) if A is a regular language, then kleene star of the language (A^*) is a regular language,
3) the empty set of a language A, denoted as $\{\}$ or $\{\lambda\}$ and
4) if A and B are regular languages, then $A \cup B$ the and $A \cap B$ are also regular languages.

III. Design

Fig. 2 describes the steps for structural vetting of academic proposals. The design presents an end-to-end technique of information extraction and vetting of an academic proposal.

A. Overview

An input proposal document is parsed (see Fig. 1 for parse tree structure), by an algorithm into meaningful document symbolic parts consisting of document's sections, chapters and pages. The leaf nodes of the parse tree structure are the document's pages symbols. In our implementation, a valid document page symbol acceptance is further determined by a REGEX parser for recognising and determining the *elements* within the symbolic string. Such elements that may be contained within a proposal document's page symbolic string are document's title, document's author, supervisor names, proposal date, section title, paragraph's heading, paragraph, figure and so on. *See Fig. 3 and 4 for the description of our proposal document elements*. The elements of a proposal document as presented in these figures are the building blocks for a proposal document organization. They appear in a predefined order for an organized paper. The structure of a proposal document is accepted or rejected after full parsing of the input proposal document into parts that validates its structure (i.e. arrangement). The correctness of the elements' structure is based on the CFG Production rules defined and successful passing by REGEX matching of elements contained in the document's parts.

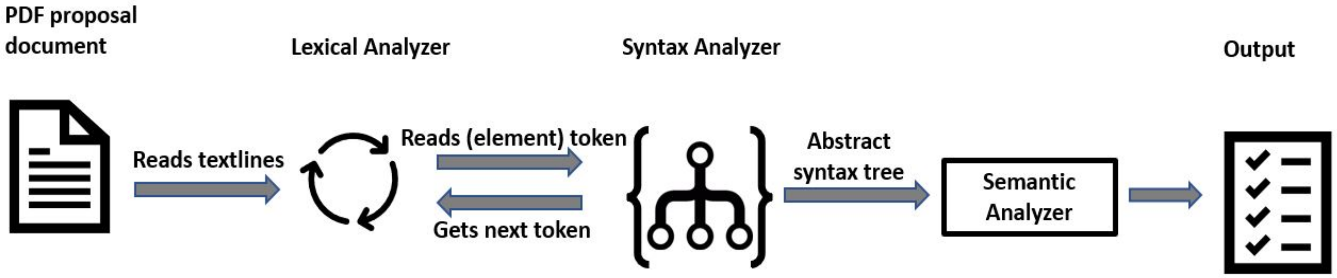


Fig. 2. Proposal Structural Vetting Design

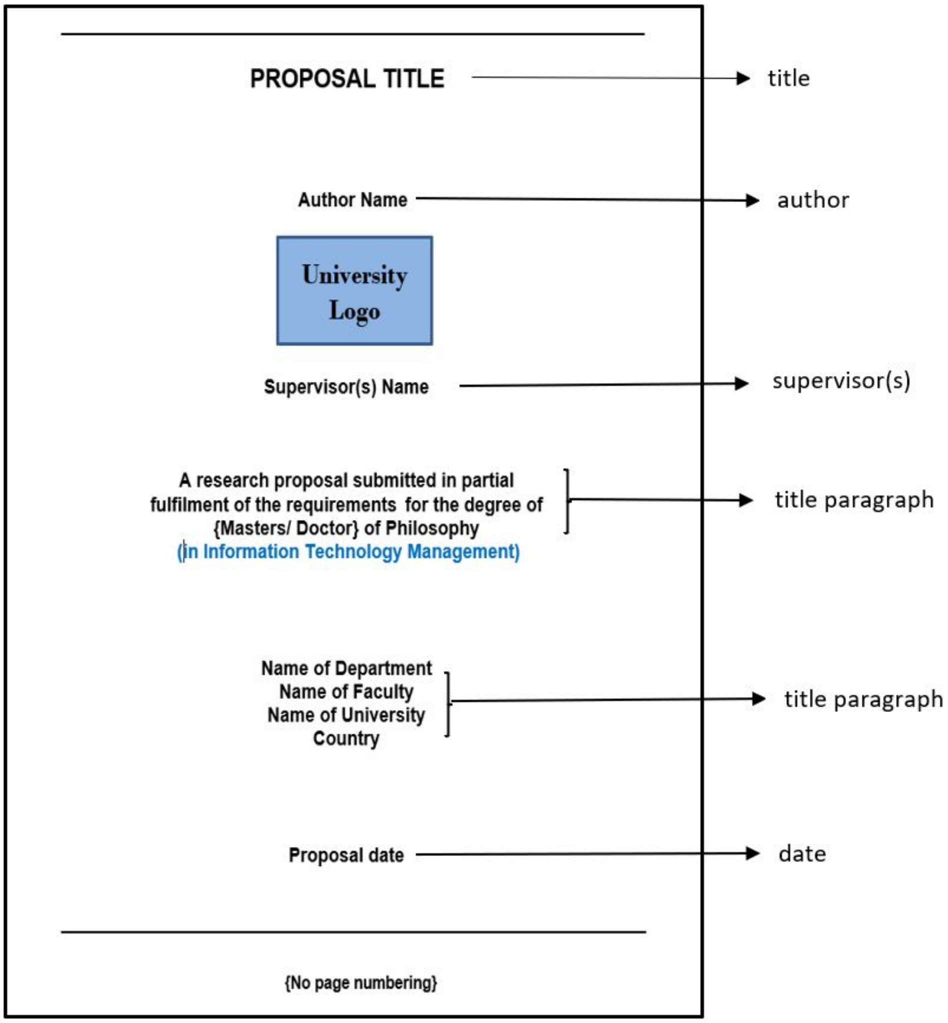


Fig. 3. Structure of Title Page

B. Lexical Analysis of PDF Proposal Document

The input document is tokenized into strings of texts identified by text line sequence on the document page. The text's feature (appearance) on the page described in the Table III is used to categorize text and "line of texts" of a page into meaningful document elements. For instance, a proposal document title is made up of line(s) of text in bold and appears on the first page of a proposal document. The input proposal is analyzed into all the elements that may be contained in a document and as described in Fig. 3 and 4.

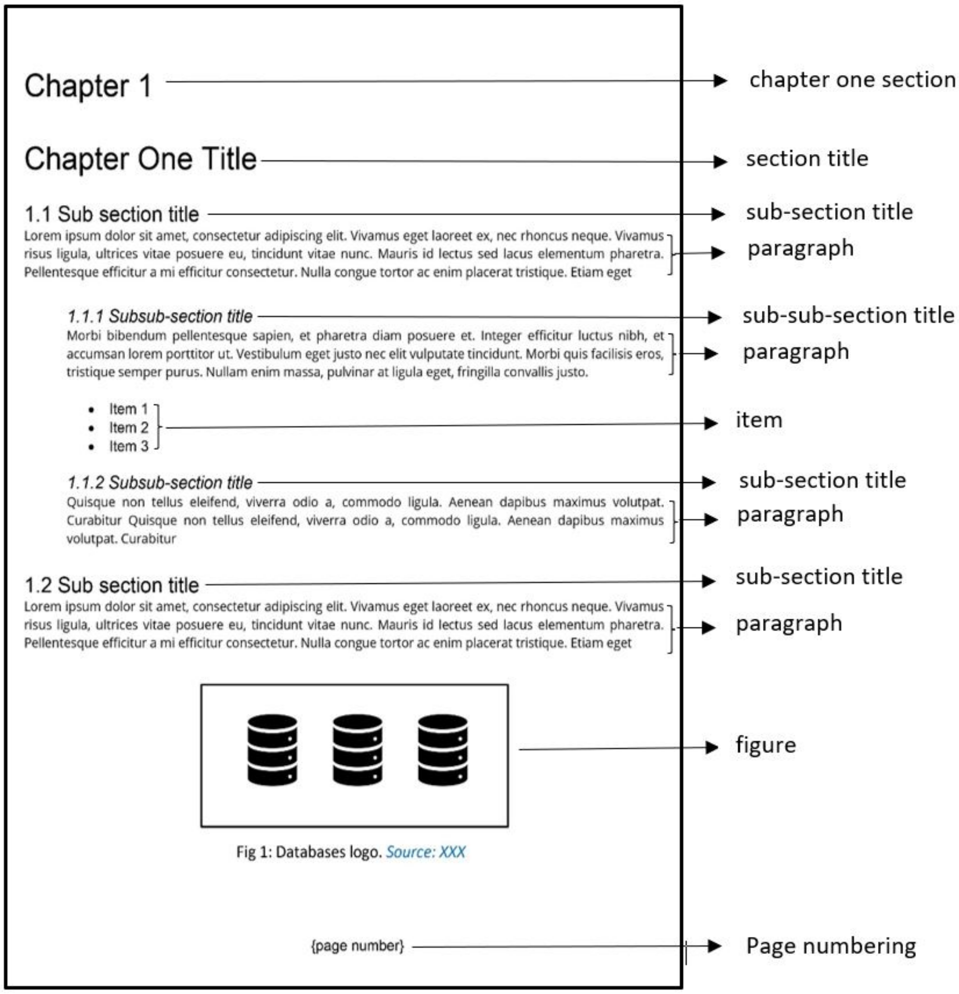

Chapter 1
chapter one section
Chapter One Title
section title
1.1 Sub section title
sub-section title
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus eget laoreet ex, nec rhoncus neque. Vivamus risus ligula, ultrices vitae posuere eu, tincidunt vitae nunc. Mauris id lectus sed lacus elementum pharetra. Pellentesque efficitur a mi efficitur consectetur. Nulla congue tortor ac enim placerat tristique. Etiam eget
paragraph
1.1.1 Subsub-section title
sub-sub-section title
Morbi bibendum pellentesque sapien, et pharetra diam posuere et. Integer efficitur luctus nibh, et accumsan lorem porttitor ut. Vestibulum eget justo nec elit vulputate tincidunt. Morbi quis facilisis eros, tristique semper purus. Nullam enim massa, pulvinar at ligula eget, fringilla convallis justo.
paragraph
Item 1
Item 2
Item 3
item
1.1.2 Subsub-section title
sub-section title
Quisque non tellus eleifend, viverra odio a, commodo ligula. Aenean dapibus maximus volutpat. Curabitur Quisque non tellus eleifend, viverra odio a, commodo ligula. Aenean dapibus maximus volutpat. Curabitur
paragraph
1.2 Sub section title
sub-section title
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus eget laoreet ex, nec rhoncus neque. Vivamus risus ligula, ultrices vitae posuere eu, tincidunt vitae nunc. Mauris id lectus sed lacus elementum pharetra. Pellentesque efficitur a mi efficitur consectetur. Nulla congue tortor ac enim placerat tristique. Etiam eget
paragraph
figure
Fig 1: Databases logo. Source: XXX
{page number}
Page numbering

Fig. 4. Structure of Chapter Pages

TABLE III. Document Component/Element List of Characteristics

Characteristics	Description
Font bold	the thickness of the text
Font size	text height
Line top spacing	height of space between the current line and the previous line.
Line bottom spacing	height of space between the current line and the next line.
Line left align	width of space between the document layout left indent and the line left co-ordinate position.
Line right align	width of space between the document layout right indent and the line right co-ordinate position
Line align type	determined as left, center or right based on the left and right indent

C. Description of Proposal Document Abstract Parse Tree using CFG

To explain the parse tree structure and its nodes, we implement a CFG with a four-tuple as given below:

$$G = (N, \Sigma, P, S) \tag{1}$$

with the following representations:

1) Set of non-terminal variables (N): a collection of all the non-terminal symbols for the production which represents proposal document's section. The symbols are described as Preliminary Section (S_p), Chapter Section (S_c), Appendix-Section (S_a) and References-Section(S_r). $N = \{S_p, S_c, S_a, S_r\}$.
2) Finite set of terminal symbols (Σ): a set of all symbols representing the proposal document terminals. These terminals are symbolic of proposal document pages and described as Title Page (t_p), Declaration Page (d_p), Acknowledgment Page (a_p), Content Page (c_p), List of Figures (l_f), List of Tables (l_t), Abstract Page (a_{bp}), Introduction Chapter Page (c_{intro}), Literature Chapter Page (c_{lit}), Methodology Chapter Page (c_{method}), Expected Contribution Page (c_{contr}), Workplan Chapter Page (c_{plan}), Conclusion Chapter Page (c_{concl}), Appendix Page (p_a) and Reference Page (p_r)

3) Set of Productions (P): are the Production rules (P) which defines the structure of the elements in the pages of the proposal document.

$$S_{doc} \longrightarrow S_p S_c S_a S_r \tag{2}$$

$$S_p \longrightarrow t_p(d_p|\lambda)(a_p|\lambda)c_p(l_f|\lambda)(l_t|\lambda)a_{bp} \tag{3}$$

$$S_c \longrightarrow c_{intro}{\cdot}c_{lit}{\cdot}c_{method}{\cdot}(c_{contr}|\lambda){\cdot}(c_{plan}|\lambda){\cdot}c_{concl} \tag{4}$$

$$S_a \longrightarrow p_a|p_a S_a|\lambda \tag{5}$$

$$S_r \longrightarrow p_r \tag{6}$$

4) Non-terminal start symbol (S): The start symbol for this grammar is the input document proposal, denoted by symbol S_{doc} .

D. REGEX for Terminal Symbols Matching

The terminal symbols of the CFG tuple are further broken down into regular expressions for recognising the page elements implemented by the symbols. The terminal symbols can be substituted with the following symbols in REGEX below.

$$t_p = e_t e_a e_s e_{tp}^{1,2} e_d \tag{7}$$

$$d_p = e_{st} e_p^{+} e_n \tag{8}$$

$$a_p = e_{st} e_p^{+} e_n \tag{9}$$

$$c_p = e_{st} e_{ta} e_n \tag{10}$$

$$l_f = e_{st}(e_i|e_{ta})e_n \tag{11}$$

$$l_t = e_{st}(e_i|e_{ta})e_n \tag{12}$$

$$a_{bp} = e_{st} e_p^{+} e_n \tag{13}$$

$$c_{intro} = e_{st}((e_p|e_{sst}|e_f|e_i|e_{ta})^{+} e_n)^{+} \tag{14}$$

$$c_{lit} = e_{st}((e_p|e_{sst}|e_f|e_i|e_{ta})^{+} e_n)^{+} \tag{15}$$

$$c_{method} = e_{st}((e_p|e_{sst}|e_f|e_i|e_{ta})^{+} e_n)^{+} \tag{16}$$

$$c_{contr} = e_{st}((e_p|e_{sst}|e_f|e_i|e_{ta})^{+} e_n)^{+} \tag{17}$$

$$c_{plan} = e_{st}((e_p|e_{sst}|e_f|e_i|e_{ta})^{+} e_n)^{+} \tag{18}$$

$$c_{concl} = e_{st}((e_p|e_{sst}|e_f|e_i|e_{ta})^{+} e_n)^{+} \tag{19}$$

$$p_a = e_{st} e_f^{+} e_n \tag{20}$$

$$p_r = e_{st} e_r^{+} e_n \tag{21}$$

TABLE IV. REGULAR EXPRESSIONS FOR DOCUMENT ELEMENTS PATTERN MATCHING

SN	Document Elements	Description	Example
1.	Title (e_t)	(\w\s)$^{+}$	Structural vetting …
2.	Author (e_a)	(By:|By)?(\s \w)$^{+}$	Joe Smith
3.	Supervisor (e_s)	(Supervisor:|$Supervisors$:)? (\s\w)$^{+}$	Prof Tim
4.	Title-paragraph (e_{tp})	((\w \s)$^{+}$ [\n])$^{+}$	This research paper …
5.	Section-title (e_{st})	(\s \w)$^{+}$	Introduction
6.	Sub-section-title (e_{sst})	\d^{+}.\d^{+} (\s \w)$^{+}$	1.1 Background
7.	Paragraph (e_p)	((\w \s)$^{+}$ [\n])$^{+}$	The advent of ICT …
8.	Page Number (e_n)	\d^{+}	2
9.	Date (e_d)	{Date API}	3rd Dec 2021
10.	Ref item (e_r)	{Reference API}	K.Van (2020), “The choice” …
11.	Figure (e_f)	{Image}	Image
12.	Item list (e_i)	((\w \s)$^{+}$[\n])$^{+}$	1. Name 2. Subject
13.	Table (e_{ta})	((\w \s)$^{+}$[\n])$^{+}$	Table

The symbols \w, \s, \n, \d matches single word, single space, a newline character and single digit respectively

E. REGEX for Document Elements Matching

The elements contained in a proposal document's pages can be matched with corresponding regular expression defined in the Table IV. The regular expressions stated are used to represent the language of the structure in English alphabet strings for every element that may be contained in a proposal document. We define a language acceptor algorithm which accepts or rejects the element token based on the matching regular expression. The language acceptor determines if the element is well-formed and which element class it belongs to. The proposal document is vetted structurally correct if all the parsed elements are accepted by the language acceptor.

F. Document Parsing

This section explains possible derivations of a proposal document given the grammar defined and its set of Production rules. Given the input proposal document which is of the start symbol S as given below, then the document can be parsed as follows:

$$S_{doc} \Longrightarrow S_p \cdot S_c \cdot S_a \cdot S_r \quad (rule\ 2) \tag{22}$$

A proposal document can be made up of the following section parts: Preliminary Section(S_p), Chapter Section(S_c), Appendix Section(S_a), Reference Section(S_r).

$$\Longrightarrow t_p \cdot d_p \cdot c_p \cdot a_{bp} \cdot S_c \cdot S_a \cdot S_r \quad (rule\ 3) \tag{23}$$

The Preliminary Section can be made up of Title Page (t_p), Declaration Page (d_p), Contents Page (c_p) and Abstract Page (a_{bp})

$$\Longrightarrow t_p{\cdot}d_p{\cdot}c_p{\cdot}a_{bp}{\cdot}c_{intro}{\cdot}c_{lit}{\cdot}c_{method}{\cdot}c_{plan}{\cdot}c_{concl}{\cdot}S_a{\cdot}S_r \ (rule\ 4) \tag{24}$$

The Chapter Section can be made up of Introduction Chapter (c_{intro}), Literature Review Chapter (c_{lit}), Methodology Chapter (c_{method}), Workplan Chapter (c_{plan}) and the Conclusion Chapter (c_{concl}).

$$\Longrightarrow t_p{\cdot}d_p{\cdot}c_p{\cdot}a_{bp}{\cdot}c_{intro}{\cdot}c_{lit}{\cdot}c_{method}{\cdot}c_{plan}{\cdot}c_{concl}{\cdot}S_r \quad (rule\ 5) \tag{25}$$

The Proposal document may not contain an Appendix Section.

$$\Longrightarrow t_p{\cdot}d_p{\cdot}c_p{\cdot}a_{bp}{\cdot}c_{intro}{\cdot}c_{lit}{\cdot}c_{method}{\cdot}c_{plan}{\cdot}c_{concl}{\cdot}p_r \quad (rule\ 6) \tag{26}$$

The Derivation 26 is an instance of a valid proposal structure and has been successfully parsed by the four-tuple grammar defined for the structural vetting of a proposal document. The expressions contained in Derivations 22 to 26 shows the complete parsing. Further understanding of the document's structure is done by recognising the document *elements* contained in each symbolic string of the language generated by the parser, refer to Derivations 7 to 21. Elements are also matched by their corresponding REGEX description in Table IV

G. Algorithms

This algorithm for the automatic structural vetting of an academic proposal document takes as input, the proposal document, and outputs the document with highlights of any structural defects that may have been picked. We present the algorithm as given below:

Algorithm 1: Structural Vetting Algorithm

```
Data: PDF proposal document, $S_{doc}$
Result: Highlighted PDF proposal document, $S_{hl}$
Parse input document into document parts using
  PDFLibrary.Extract($S_{doc}$)
if ValidParse(Document parts) then
    if REGEX_Match(Elements contained in
      Document Parts) then
        if REGEX_Match(words contained in
          Document Elements) then
            PROCESS COMPLETE: Document
              verified successful
            Display success message
        end
        else
            ERROR DETECTED: Highlight error on
              Proposal
        end
    end
    else
        ERROR DETECTED: Highlight error on
          Proposal
    end
end
else
    ERROR DETECTED: Highlight error on Proposal
end
Document_file_location $\longleftarrow$
  Save_Highlighted_Document_toDisk($S_{hl}$)
return Return $S_{hl}$
```

IV. Implementation and Results

The algorithms presented in this paper were implemented using C sharp Microsoft's .Net library. The software embeds an iTextSharp version 7.0 library used for PDF document tokenization and manipulation of text. The Auto-proofreader tool interface displaying its menu controls is presented in Fig. 5.

Table V shows the results from the implementation of a corpus of 20 academic proposals. All the results vary in the degree of the number of error or success feedback reported by the tool. All the elements in each document have been parsed against defined rules and REGEX patterns based on a given instruction format guide.

A. Description of Dataset

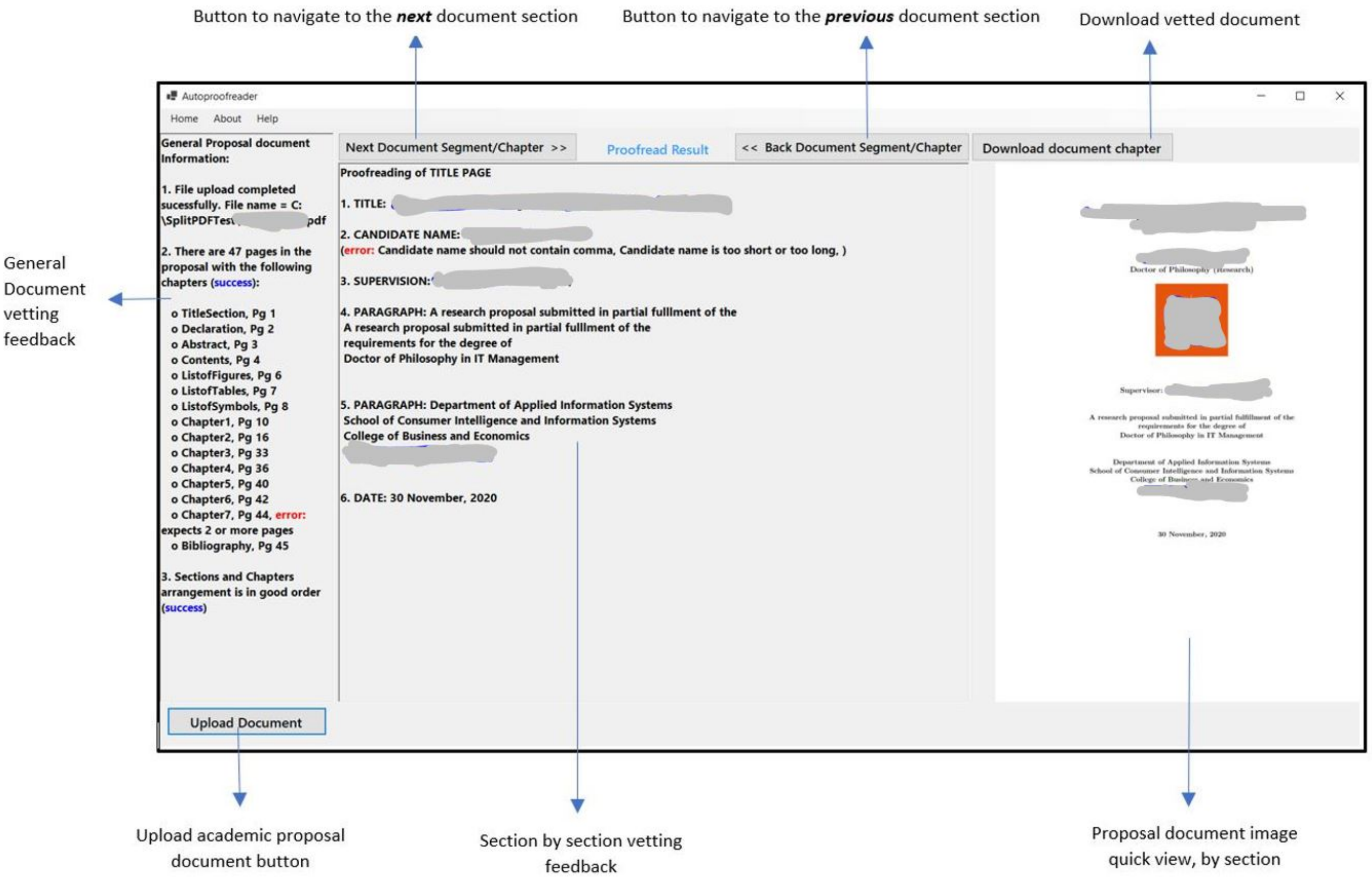
A corpus of ten (20) proposals collected from the online digital ecommons of various Universities was used to carry out the experimental result. The dataset consists of academic proposals of MSc and PhD post-graduate students in Information Systems' related discipline that were accessed on Universities institutional repository (IR) or available on online digital-common. These are academic proposal thesis submitted between year 2015 and 2021.

TABLE V. Results of Vetting of 20 Academic Proposals

S/N	Document Elements	TP	TN	FP	FN	Prec	Rec	Acc
1	Title	10	5	5	0	0.67	1.00	0.75
2	Author	5	5	9	1	0.36	0.83	0.50
3	Supervisor	15	5	0	0	1.00	1.00	1.00
4	Title-Paragraph	21	16	0	3	1.00	0.88	0.93
5	Section-Title	111	0	0	13	1.00	0.90	0.90
6	Sub-Section	2041	50	0	826	1.00	0.71	0.72
7	Paragraph	4019	255	0	538	1.00	0.88	0.88
8	Page-No	3011	0	0	0	1.00	1.00	1.00
9	Date	6	1	0	13	1.00	0.32	0.35
	Total	9239	337	14	1394	0.99	0.87	0.87

B. Result and Discussion

The difference in the number of correctly identified document elements by the tool differs significantly for some logical elements due to the complexities of computationally identifying them. The tool detected and proofread the structural items on the preliminary pages with better precision and accuracy. The overall tool precision and accuracy is given as 0.99 and 0.87, respectively. Refer to expressions given in Equation 27 and 29


Button to navigate to the *next* document section
Button to navigate to the *previous* document section
Download vetted document
General Document vetting feedback
Autoproofreader
Home
About
Help
Next Document Segment/Chapter >>
Proofread Result
<< Back Document Segment/Chapter
Download document chapter
Upload Document
Upload academic proposal document button
Section by section vetting feedback
Proposal document image quick view, by section

General Proposal document Information:

1. File upload completed sucessfully. File name = C: \SplitPDFTes[redacted]pdf

2. There are 47 pages in the proposal with the following chapters (success):

- TitleSection, Pg 1
- Declaration, Pg 2
- Abstract, Pg 3
- Contents, Pg 4
- ListofFigures, Pg 6
- ListofTables, Pg 7
- ListofSymbols, Pg 8
- Chapter1, Pg 10
- Chapter2, Pg 16
- Chapter3, Pg 33
- Chapter4, Pg 36
- Chapter5, Pg 40
- Chapter6, Pg 42
- Chapter7, Pg 44, error: expects 2 or more pages
- Bibliography, Pg 45

3. Sections and Chapters arrangement is in good order (success)

Proofreading of TITLE PAGE

1. TITLE: [redacted]

2. CANDIDATE NAME: [redacted]
(error: Candidate name should not contain comma, Candidate name is too short or too long,)

3. SUPERVISION: [redacted]

4. PARAGRAPH: A research proposal submitted in partial fullment of the
A research proposal submitted in partial fullment of the
requirements for the degree of
Doctor of Philosophy in IT Management

5. PARAGRAPH: Department of Applied Information Systems
School of Consumer Intelligence and Information Systems
College of Business and Economics
[redacted]

6. DATE: 30 November, 2020

[redacted]

[redacted]
Doctor of Philosophy (Research)

[redacted]

Supervisor: [redacted]

A research proposal submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in IT Management

Department of Applied Information Systems
School of Consumer Intelligence and Information Systems
College of Business and Economics
[redacted]

30 November, 2020

Fig. 5. Autoproofreader Tool and Menus.

V. Evaluation

We present the evaluation for the tool in terms of accuracy for correctly classifying/ recognising syntactically right or wrong elements in the document. We present the confusion matrix model evaluation with the performance metrics of sensitivity, precision and accuracy given below. Table VI gives a brief explanation of how the confusion matrix is applied for evaluating the tool.

Overall Precision:

$$\frac{TP}{(TP+FP)} = \frac{9239}{(9239+14)} \approx 0.99 \tag{27}$$

Overall Recall:

$$\frac{TP}{(TP+FN)} = \frac{9239}{(9239+1394)} \approx 0.87 \tag{28}$$

Overall Accuracy:

$$\begin{aligned} & \frac{(TP+TN)}{(TP+TN+FP+FN)} \\ & = \frac{(9239+337)}{(9239+337+14+1394)} \approx 0.87 \end{aligned} \tag{29}$$

where:
TP = True Positives

[h!]

TABLE VI. Confusion Matrix Evaluation

Item	Auto-proofreader Tool(T/F)	Benchmark (P/N)	Description
TP	T = Identified	P = Correct Elements	Identified correct Elements
TN	T = Identified	N = Incorrect Elements	Identified incorrect Elements
FP	F = Did not identify	P = Correct Elements	Did not identified correct Elements
FN	F = Did not identify	N = Incorrect Elements	Did not identified incorrect Elements

TN = True Negatives
FP = False Positives
FN = False Negatives

VI. Conclusion and Future Work

Identifying errors in a (lengthy) textual document is an expert task that can be made into an artificial intelligent (AI) tool which can then assist human reviewers (i.e. Supervisors) to more effective and productive, especially when faced with many documents (i.e. academic proposals). The tool will not only be useful for academic proposals but can be refined for many varying template-based documents. In this paper, we have presented the technique and design of the tool for vetting an academic proposal for layout or structure-based errors. This technique slices a proposal document into its minute structural components (which we have termed as document elements contained a document section) and performs a check of correctness. The technique allows the slicing of the academic proposals into separate sections as documents and lastly allows the download of a **vetted** academic proposal document which can be used as a feedback to Students' candidate after an automatic vetting. The designed software tool was evaluated on twenty (20) proposal documents which gives an accuracy of 86%. The accuracy of the tool is only based on specific elements for evaluation but it can be made more robust with vetting more components (or elements) contained in a document section such as images (or figures) and tables. The software design is template specific but can be extended for various kind of template-driven text document vetting.

In the future, we will explore the structural vetting of more document logical elements: figures/images, tables and citations. We will also extend the tool to perform rule-based sentence level grammar vetting.

Acknowledgment

This research was partially supported by the Global Excellence Stature (GES) awards and National Research Fund (NRF) with grant number - 119041

References

[1] S. P. Glenda Rosy Clements, "Application of natural language processing in document vetting," *PSYCHOLOGY AND EDUCATION*, vol. 57, pp. 5651–5658, 11 2020.

[2] H. Zhong, C. Xiao, C. Tu, T. Zhang, Z. Liu, and M. Sun, "How does nlp benefit legal system: A summary of legal artificial intelligence," 2020.

[3] R. Dale, "Law and word order: Nlp in legal tech," *Natural Language Engineering*, vol. 25, pp. 211–217, 01 2019.

[4] O. Daramola, "Lessons from postgraduate supervision in two african universities: An autoethnographic account," *Education Sciences*, vol. 11, p. 345, 07 2021.

[5] G. van Rensburg, P. Mayers, and L. Roets, "Supervision of postgraduate students in higher education," *Trends in Nursing*, vol. 3, 11 2016.

[6] N. S. Sudheesh K, Duggappa Duggappa Rani, "How to write a research proposal," pp. 631–634, 09 2016.

[7] M. Bushesha, H. Mtae, J. Msindai, and S. Mbogo, "Challenges facing supervisors and students in the process of writing theses/dissertations under odl: Experiences from the open university of tanzania," *Huria: Journal of the Open University of Tanzania*, vol. 12, pp. 118–131, 2012.

[8] G. van Rensburg, P. Mayers, and L. Roets, "Supervision of postgraduate students in higher education," *Trends in Nursing*, vol. 3, 11 2016.

[9] M. d. K. Jan Botha, Gabriele Beata Vilyte. (2019). [Online]. Available: "https://theconversation.com/digital-training-can-help-supervisors-lift-phd-output-126391"

[10] A. J. Hey-Cunningham, M.-H. Ward, and E. J. Miller, "Making the most of feedback for academic writing development in postgraduate research: Pilot of a combined programme for students and supervisors," *Innovations in Education and Teaching International*, vol. 58, no. 2, pp. 182–194, 2021. [Online]. Available: https://doi.org/10.1080/14703297.2020.1714472

[11] K. Wilmot and H. Lotz-Sisitka, *Supporting Academic Writing Practices in Postgraduate Studies. A sourcebook of academic writing support approaches and initiatives*, 2015.

[12] H. D. Kohn, "A mentoring program to help junior faculty members achieve scholarship success," *American Journal of Pharmaceutical Education*, vol. 78, 2014.

[13] N. Bak, *Research Proposal Guide*, 04 2015.

[14] L. Karyuatry and M. Rizqan, "Grammarly as a tool to improve students' writing quality: Free online-proofreader across the boundaries," *JSSH (Jurnal Sains Sosial dan Humaniora)*, vol. 2, p. 83, 05 2018.

[15] K. Siau and Y. Ma, "Artificial intelligence impacts on higher education," 05 2018.

[16] R. F. Heller, *The Distributed University for Sustainable Higher Education*. Springer, 2022.

[17] D. Maor and J. K. Currie, "The use of technology in postgraduate supervision pedagogy in two australian universities," *International Journal of Educational Technology in Higher Education*, vol. 14, no. 1, pp. 1–15, 2017.

[18] J. Miranda, C. Navarrete, J. Noguez, J.-M. Molina-Espinosa, M.-S. Ramírez-Montoya, S. A. Navarro-Tuch, M.-R. Bustamante-Bello, J.-B. Rosas-Fernández, and A. Molina, "The core components of education 4.0 in higher education: Three case studies in engineering education," *Computers & Electrical Engineering*, vol. 93, p. 107278, 2021.

[19] P. Race, "Some pros and cons of 'track changes' feedback on work returned to students electronically," 03 2014.

[20] B. P. Rasmus, "End to end information extraction from business documents," 2019.

[21] S. Abbott and A. Ade-Ibijola, "Algorithms and a tool for automatic decryption of clinical notes," *2019 6th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, pp. 137–143, 2019.

[22] S. Kubeka and A. Ade-ibijola, "Automatic Comprehension and Summarisation of Legal Contracts," *Advances in Science, Technology and Engineering Systems Journal*, vol. 6, no. 2, pp. 19–28, 2021.

[23] S. Singh, "Natural language processing for information extraction," 2018.

[24] H. Jiang, Y. Hua, D. Beeferman, and D. Roy, "Annotating the tweebank corpus on named entity recognition and building nlp models for social media analysis," *arXiv preprint arXiv:2201.07281*, 2022.

[25] G. Obaido, A. Ade-Ibijola, and H. Vadapalli, "Talksql: A tool for the synthesis of sql queries from verbal specifications," *2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, pp. 1–10, 2020.

[26] T. Revanth, K. V. Sai, R. Ramya, R. Chava, V. Sushma, and B. Ramya, "Nl2sql: Natural language to sql query translator," pp. 267–278, 2022.

[27] A. Gheewala, C. Turner, and J.-R. de Maistre, "Automatic extraction of legal citations using natural language processing," in *Proceedings of the 15th International Conference on Web Information Systems and Technologies*, ser. WEBIST 2019. Setubal, PRT: SCITEPRESS - Science and Technology Publications, Lda, 2019, p. 202–209. [Online]. Available: https://doi.org/10.5220/0008052702020209

[28] M. Cronje and A. Ade-Ibijola, "Automatic slicing and comprehension of cvs," in *2018 5th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. IEEE, 2018, pp. 99–103.

[29] K. Ahmad, M. A. Ayub, K. Ahmad, J. Khan, N. Ahmad, and A. Al-Fuqaha, "Merit-based fusion of nlp techniques for instant feedback on water quality from twitter text," *arXiv preprint arXiv:2202.04462*, 2022.

[30] A. Alamoudi, A. Alomari, S. Alwarthan, and A. Rahman, "A rule-based information extraction approach for extracting metadata from pdf books," *ICIC Express Letters*, vol. 12, pp. 121–132, 02 2021.

[31] H. Assal, J. Seng, F. Kurfess, E. Schwarz, and K. Pohl, "Semantically-enhanced information extraction," in *2011 Aerospace Conference*, 2011, pp. 1–14.

[32] K. Verma, A. Kass, and R. Vasquez, "Using syntactic and semantic analyses to improve the quality of requirements documentation," *Semantic Web*, vol. 5, pp. 405–419, 01 2014.

[33] A. Ade-Ibijola, "Finchan a grammar-based tool for automatic comprehension of financial instant messages," *Proceedings of the Annual Conference of the South African Institute of Computer Scientist and Information Technologists*, vol. 0, no. 1, p. 0, 2016.

[34] X. Chen, H. Xie, G. Cheng, L. Poon, M. Leng, and F. L. Wang, "Trends and features of the applications of natural language processing techniques for clinical trials text analysis," *Applied Sciences*, vol. 10, p. 2157, 03 2020.

[35] E. S. Chifu, V. R. Chifu, I. Popa, and I. Salomie, "A system for detecting professional skills from resumes written in natural language," in *2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2017, pp. 189–196.

[36] A. Anjewierden, "Aidas: incremental logical structure discovery in pdf documents," *Proceedings of Sixth International Conference on Document Analysis and Recognition*, pp. 374–378, 2001.

[37] H. Langer and P. Bayerl, "Text type structure and logical document structure," 07 2004.

[38] S. O. M. Perdana, Indra and F. A. Masri, "Effectiveness of online grammarly application in improving academic writing: Review of experts experience." *International Journal of Social Sciences*, 2021.

[39] M. Nova, "Utilizing grammarly in evaluating academic writing: A narrative research on efl students' experience," *Premise: Journal of English Education*, vol. 7, p. 80, 04 2018.

[40] L. Karyuatry and M. Rizqan, "Grammarly as a tool to improve students' writing quality: Free online-proofreader across the boundaries," *JSSH (Jurnal Sains Sosial dan Humaniora)*, vol. 2, p. 83, 05 2018.

[41] T. Nur Fitria, ""grammarly" as ai-powered english writing assistant: Students' alternative for english writing," *Metathesis Journal of English Language Literature and Teaching*, vol. 5, pp. 65–78, 05 2021.

[42] A. Nasution and S. Fatimah, "The use of pro writing aid web in editing students writing," *Journal of English Language Teaching*, vol. 7, no. 2, pp. 362–368, 2018.

[43] Y. Luo, J. Yu, and X. Cheng, "The research of chinese text proofreading system model," 12 2020.

[44] A. Ade-Ibijola, "Synthesis of regular expression problems and solutions," *International Journal of Computers and Applications*, vol. 42, no. 8, pp. 748–764, 2020. [Online]. Available: https://doi.org/10.1080/1206212X.2018.1482398

[45] J. C. Martin, *Introduction to languages and the theory of computation, fourth edition*. New-York: McGraw-Hill, 2010.

Mobile Application Prototype: Learning in the Programming Course in Computer Engineering Students

Lilian Ocares-Cunyarachi, Laberiano Andrade-Arenas
Facultad de Ciencias e Ingeniería
Universidad de Ciencias y Humanidades
Lima, Perú

Abstract—**Students need to continue with the learning process related to the world of programming because today are in the era of technological globalization. Therefore, it is very important to learn about it, since programming is used in different areas and as a result obtain software, electronic devices, among others. seek to design a mobile application that helps students learn much more about programming, since students in the first cycles of computer science and computer science have difficulties learning about different programming languages. That is why the application seeks to help the student by complementing their learning in such a way that they can obtain favorable results in their progress thanks to the development of the application. The objective is to design a mobile application for teaching programming in a didactic way that helps computer science students with learning difficulties. The methodology used is Design Thinking, because it is an agile methodology that is based on phases that help us understand and collect information about the problem encountered in order to provide a solution. As for the case study, the design of the mobile application and the detailed development of the prototype are shown. The result obtained is the prototype of the mobile application in which students with learning difficulties will benefit. In addition, a survey carried out at the University of Sciences and Humanities to students and teachers is shown, where very relevant data is obtained according to their learning.**

Keywords—Design thinking; learning; mobile application; students; programming.

I. Introduction

Due to The worldwide pandemic by the coronavirus or also called Covid-19 has led to a global rethinking as teaching either in schools or universities, around the world countries had to close student centers for several weeks due to the pandemic, therefore, technology has taken huge steps in a very surprising and significant way [1],[2],[3]. An important factor is programming; because websites, applications, software and everything have as a tool today to work from home or remotely or take classes from the comfort of home requires devices that are developed based on codes and that is where programming becomes of paramount importance worldwide [4],[5].

The education related to programming is very essential for the technological development in Latin America and around the world, this is because most of the organizations, in the labor and student field are subject to the applications that are created through codes coded in different programming languages performing their functionalities correctly [6]. However, programming is involved in the development of different industries [7]. Nowadays learning the art of programming has a high transcendence in the trade field as learning English, since nowadays it is very essential worldwide, therefore programming can not only be useful for computer engineers or computer technicians, because programming is present in everything that surrounds us, that is why it can be for everyone who has a university degree as well as for those who do not have one, it does not vary in the result of learning. It should be noted that the most interesting thing about a programmer is that he can create new things from 0 [8] by means of codes for the resolution of existing problems.

Unfortunately, the Latin American continent does not have enough experts to be a world power in the field of software or information technology in general [9]. Due to the fact that there is no incentive since childhood in the art of programming, so it is important to encourage children today as well as young people with seminars or courses to join their own research establishments related to computer science [10]. Being more and more innovative, having as direct consequence the economic and social growth of the country where an innovative method is applied in such a way to begin to create a better future [11],[12].

Several university students of the first cycles in the University of Sciences and Humanities return to take the course of Programming is why the importance of this research is to help students with this mobile application so they can learn more about programming in a didactic way since that is an advantage for students in the student subject. Likewise, Peruvian universities in an applicative approach in programming generate great uncertainty to their students of the first cycles, since, they only count on referring their grades. The most common problems in these students is the difficulty of learning programming languages. For all these reasons, the present investigation took as a reference the University of Sciences and Humanities in the faculty of computer engineering where students of the first three semesters are involved, identifying some problems of complication in programming.

The methodology to be implemented is Design Thinking, which is an innovative procedure to produce outstanding ideas with great effectiveness in understanding and offering a solution to the needs of the users in order to obtain favorable results.

The implementation of the mobile application is very important, so formulate the following question: How will it improve the learning of computer engineering students in the programming course?

The objective of this research work is to implement a prototype of a mobile application to improve learning in a didactic way with the students of the first cycles of the programming language course.

The paper is structured as follows: Section II describes in detail the literature review. Section III shows the methodology, Section IV shows the case study, Section V shows the results and discussion and finally, Section VI presents the conclusion and Section VII presents the future work.

II. Literature Review

This section presents an overview of different studies on learning programming in a way that is easier for students to understand. Not only using different methods but also explaining why the different methods should be used creatively depending on the results.

According to the author [13], he mentions that the development of mobile applications is a group of processes and procedures that relate to software for different wireless computing devices either small or large, such as smartphones among others.

Therefore, the author [14], refers that the development of web applications and mobile applications has its origin in open source development. However, a very important difference is that mobile applications are commonly written specifically to take advantage of the unique properties or functionalities of a mobile device. For example, a gaming application can take advantage of the phone's accelerometer, just as a health device can take full advantage of a smart watch's temperature sensor.

Likewise, the authors [15], highlight that the most important mobile platforms today are Apple's iOS and Google's Android plus an important fact is that Apple phones and tablets come preinstalled with essential applications, including a full web browser and the app store. Android devices also come preinstalled with similar applications through the Google Play Store now live in an era of technology agigantada, where interact daily with our mobile devices, in this context, this research project is aimed at developing an interactive mobile application to enhance the learning process to students through programming.

Education is very important for the formation of students in general, likewise it is advisable to integrate the use of mobile applications in the teaching-learning process of all other fields, to take advantage of the benefits brought by mobile technology and promote students to create a creative study habit [16]. Concluding that the optimal results have been obtained, successfully implemented in the learning process related to programming, thus having a significant impact on this type of implementation; obtaining a comfortable result.

Most first-year computer science students will find learning object-oriented programming difficult. Serious games have once been used as an approach to handle this problem. But most of them cannot be played on mobile devices. Obviously, this does not suit the era of mobile computing that aims to enable students to learn programming skills anytime, anywhere, thereby enhancing the learning of programming languages. A research project started more than a year ago and aims to create a serious gaming approach related to mobile devices along with an educational game for learning programming To date, the project has conducted a literature review to understand the existing work and identify problems in the field, conducted surveys to find out the needs of students for a mobile device based approach and then set up a serious mobile device based gaming approach with a developed prototype of the game [17]. It is expected that the presented project will be useful and helpful to integrate more effective approaches with didactic mobile games for learning object oriented programming in such a way to enhance the learning experience for university students.

Thus, the authors [18], describe that mobile technologies have a great impact on education and teaching of computer science programming, which leads to the development of tools to benefit this process in introductory courses. The increased use of cell phones makes it possible to encourage their use in programming courses using a mobile platform, seeking to improve classes with out-of-class support. It shows the structure and experience of use of Paepoo, Platform for Learning and Education of Object-Driven Programming, a mobile application where students, enthusiastic about the use of their cell phones, accept an active and committed role in their learning process, taking advantage of the support platform with great results and approval.

The authors [19], emphasize that to explicitly model complex dependencies between applications, adopt a dynamic graph structure to learn users' interests. First, extracted users' interests in each application usage graph using the hierarchical graph attention mechanism. Second, capture the time evolving user interests and generate the dynamic user embeddings by modeling the temporal dependencies among multiple application usage graphs. Finally, obtain the current user interests in the current application usage graph, merge the interests of multiple users, and generate comprehensive user embeddings for the next mobile application recommendation.

The authors [20], mention the rapid development of information technologies makes it possible to create and innovate more mobile devices. Most of the distance learning students require access to analysis materials such as communication tools and also extra learning media not only at home or at their workplace. The purpose of this article is to expose the modalities of mobile technology in computer science and programming education. According to the results of surveys conducted in elementary schools, high schools and universities. It is possible to mention that mobile devices are used more and more in relation to learning. The results of surveys and experiments show that mobile devices have the potential to improve education in computer science, programming and algorithms. The article explains the experience of teaching and development of mobile applications, for teaching and for users with special needs.

Likewise, the authors [21], researched and analyzed the development of a mobile application for the education of basic concepts such as programming. The purpose is to help students acquire skills while having fun and using their own devices. The mobile application was designed according to a cross-platform approach to reach the widest possible audience of students, saving development and maintenance time and effort. The code is fully shared between IOS, Android and Windows mobile platforms, allowing students to install the app on any device. The core application is based on a multivalent system to make the app interactive, flexible and dynamic and provide students with personalized instructions [22]. A prototype showing the main features of the application is presented.

In summary, the authors analyzed in their research conclude that over the years in the field of learning together with technology has been advancing rapidly so must adapt to the era of globalization using beneficial tools that help us to be able to do new things from 0 as is the programming applied in computer science.

III. Methodology

The Design Thinking process is made up of 5 phases and the Design Thinking methodology contains iterative processes, it is not linear, therefore it is a process that serves to address complicated challenges made up of the so called wicked problems or drawbacks, drawbacks that are complicated to conceptualize and solve as they was discovered during the practice of the process of the methodology. The most interesting thing is that at any time you can take steps forward or backward if required in the Design Thinking process, jumping even to non consecutive stages, collecting information and generating a huge proportion of content, which will grow or shrink depending on the stage in which are. Fig.1 shows the phases that implement next [23], by means of questions so that these are answered in an effective way in order to have good results.

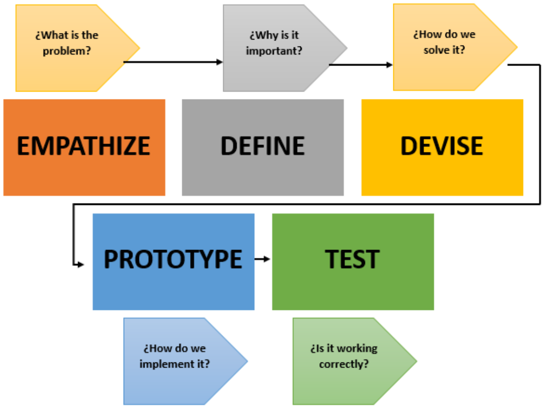


Fig. 1. Design Thinking.

Phases of Design Thinking

1) Empathy: As the first phase begins with understanding the needs, based on the requirements of the users, that is why in this first phase we must put ourselves in the situation of the users, so that from this we are able to create resolutions for a better lifestyle.

2) Define: As a second phase must process the information collected throughout the previous stage in such a way that are left with what really adds value to the research, therefore, it leads us to the scope of the perspectives. Therefore, identify drawbacks to solve the problem as a key point to obtain a satisfactory result, but above all innovative.

3) Ideation: As the third stage of Ideation, its purpose is to generate an infinite number of possibilities for the resolution of the problem encountered, which is why should not just go with the first thing that comes to mind, must analyze the situation well. At this stage, the requirements favor the creation of the application. Sometimes, the most extravagant ideas are the ones that produce creative resolutions.

4) Prototyping: As a fourth phase implement prototyping in which turn ideas into reality. That is, building prototypes directed towards the ideas helps us to visualize the probable resolutions. Therefore, at this stage can see the resources that must improve or modify in order to reach the final result.

5) Testing: Throughout the Testing stage, the prototypes will be tested with users based on the solution being implemented, therefore this stage is crucial which will allow us to detect improvements as well as failures to be solved, throughout this stage it evolves because it starts with the initiative until get to turn it into the solution were trying to find in such a way that return a properly developed final product.

IV. Case Study

1) Empathy: Applying the first stage collected detailed information by means of surveys with university students of the first cycles, in the first instance, in order to know if they had an elementary idea of what are the algorithms related to programming, therefore it was mentioned if they would use a didactic application to complement their learning in their classes. Likewise, in order to know and understand the inconveniences and problems that emerge in the students today and to understand their situation, in such a way to collect the case since it is of great importance for our inquiry and execution of this investigation. Tables I, II, and III were validated by expert judgment. Validation was of contents where the validation scale was low from 0 to 35%; from 35% to 70% medium; and from 70% to 100% high to medium. the evaluation criteria were relevance, coherence and clarity. Obtaining 85%, giving as approved by expert judgment. As it is observed in Table I, 3 questions were made to the students of first cycle of the career of computer science where the following questions were asked Do you like the art of programming? Do you have difficulties in understanding the different programming languages? And finally, is it difficult to learn to program? Thanks to these questions valuable information was collected.

TABLE I. Questions on the Taste of Programming

	Questions
Q1	¿Do you like the art of programming?
Q2	¿Do you have difficulties in understanding the different programming languages?
Q3	¿Do you find it difficult to learn to program?

2) Define: In this stage seek to detect a starting point based on the detailed information through the data obtained in the previous stage from the surveys, so that can propose the best solution for that need. As understand the student's need and have access to the different examples and tasks based on algorithms, as shown in Table II, 4 questions were asked based on stage where the following questions were asked: As a student, how important do you think algorithms are in the creation of something new? How do you create a programming algorithm? How do you develop a programming algorithm? What programming language do you often use? Thanks to these questions gathered important information.

TABLE II. QUESTIONS ON THE IMPORTANCE OF PROGRAMMING

	Questions
Q1	¿As a student, how important do you think algorithms are in the creation of something new?
Q2	¿How is a programming algorithm created?
Q3	¿How is a programming algorithm developed?
Q4	¿What programming language do you often use?

3) Ideation: Keeping in mind the two previous phases, which are empathizing and defining, several ideas were obtained to decide the functionality of the application. Likewise, the needs of the students will be aligned with the facilities provided by the implementation of the application. As shown in Table III, three questions were asked to students of the first cycle of computer science where the following questions were asked: How often do you use technological applications on your mobile device? Do you consider innovative the idea of creating a mobile application that helps to reinforce your knowledge related to programming? Do you think it would be helpful to use a mobile application to learn to program through algorithms? Thanks to these questions a clearer perspective was obtained.

TABLE III. QUESTIONS ON THE MOBILE APPLICATION

	Questions
Q1	¿How often do you use technological applications on your mobile device?
Q2	¿Do you consider innovative the idea of creating a mobile application that helps to reinforce your knowledge related to programming?
Q3	¿Do you think it would be helpful to use a mobile application to learn programming through algorithms?

4) Prototyping: Once the idea was clear, we designed the didactic application for learning programming based on the different programming languages, adding to this that is implemented with a virtual assistant (Chatbot) which will accompany and help the user to perform their learning correctly in such a way to obtain favorable results.

- Python: Python is a computer programming language widely used in the development world widely used in the development world for many different applications. many different applications. Widely used in scientific research and in fields ranging from finance to biomedicine, Python is from finance to biomedicine, Python is making inroads in many professions. many professions. The world ranking of the most popular and popular and popular programming languages shows that Python is among the top three Python is among the top three most preferred programming languages worldwide [24]. This is why it is important to introduce the Python language as a programming language in the computer and research and research centers in such a way that it can be applied in schools and in all in schools and in all Peruvian universities.

- Javascript: The JavaScript language was initially created for user-side programming in web browsers, nowadays its engine is included in various types of software programs, because the semantics of JavaScript are complex [25], especially thanks to its dynamic nature as the understanding and knowledge of JavaScript programs are challenging tasks for a programmer.

- Pseint: In introductory programming courses, students learn the logic for algorithm development. In these courses the usual methodology is to teach students the logic for the construction of algorithms in Spanish using paper, then in a second course a specific programming language is taught, usually C or Java. However, the syntax and instructions of the latter are in English [26]. This study analyzes the use of PSeInt as a support for teaching programming with syntax and instructions in Spanish.

- C#: The programs that are created have different types of formatting, i.e. the syntax that is used. The syntax in C# is a sequence of rules and processes that lead the composition of a procedure. These rules must be understood by the compiler that is executing the program in such a way as to produce a valid C# program, for example, they must implement how a line of C# code begins, how it ends and at what point to use, for example, quotation marks, brackets or braces [27]. The C# language excludes uppercase and lowercase letters, and the C# language is programmed in lowercase letters.

- Java: Nowadays we are in an environment of the Internet era, by means of the learning environment in this case of the university students suffering transcendental changes in their education, besides there has been a new change in the education, in which the learning by means of Internet became a tendency [28]. The java course is a must for every programmer, besides, it is worth mentioning that java is a very required platform, above all, safe and reliable for its development.

- Html: Learning Html presents challenges similar to those of learning a programming language, i.e. it is common for a beginner programmer to have a Html validator, and there are limited tools to help programmers who are just starting out in the programming world to fix bugs in their Html code. In this analysis, we used visualization techniques to show the structural and contextual information of Html code. looked at condensing and visually representing the relevant points of the Html code [29]. In other words, to allow novice programmers to obtain data about the composition of the Html code and to locate any semantic errors in corresponding syntax.

- Kotlin: Pure Kotlin code excludes the utilization of any java graphics library where it can be transpiled to JavaScript and realized on a web. Nonetheless, writing in Kotlin code will run without modification both in a web browser and also in jvm likewise being trivial in such a way precise adherence to a timely methodology is needed [30].
- Php 8: Php 8, was released at the end of 2020, being a fundamental update of one of the most famous programming languages so to speak so implemented in 4 out of 5 developed websites that use a server-side language also note that a free platform so the code is done on Linux, Windows. Php is subjectively easy to learn [31]. Many of the novel properties of PHP 8 where they make the code more efficient and accurate for a good development.

All the programming languages mentioned above are very essential to be able to program which is why these languages were chosen, because they are very required today in the labor market, also these languages will be implemented in the mobile application.

5) Testing: In this last phase verified in detail what was done, identifying each phase its good execution for the realization of the learning prototype related to the different programming languages, therefore a good performance was observed.

Fig. 2 shows the flowchart which begins by entering the application so that then be registered and then go to enter the application after the main screen where they will be attended by a virtual assistant who will welcome the registered user, for this the assistant will show the programming courses that are available for the user to choose any of them, you can also choose whether to start with the basic course, intermediate or advanced, Once the student has studied the selected programming language, he will go through a test of exercises where he will develop what he has learned. It is worth mentioning that if the student makes a mistake in the exercises, the virtual assistant will encourage him with some messages so that he keeps on trying, and if he solves the exercises correctly, he will be congratulated for what he has learned.

V. Results and Discussion

A) About the Case Study

In this section for the case study, it was carried out the design of the mobile application implemented with a chatbot for learning the programming course applied in computer science students as well as this application can be used by people who need to know more about the world of programming is why the implementation of the application is very beneficial because it helps to complement the knowledge acquired in addition to put them into practice as the system comes with a list of exercises so that they can be developed and increase the student's knowledge.

Fig. 3 shows the user's welcome and registration interface by entering their first name, last name, Gmail and password, and they can also log in directly from their Gmail email if they require it so that they can register in the application.

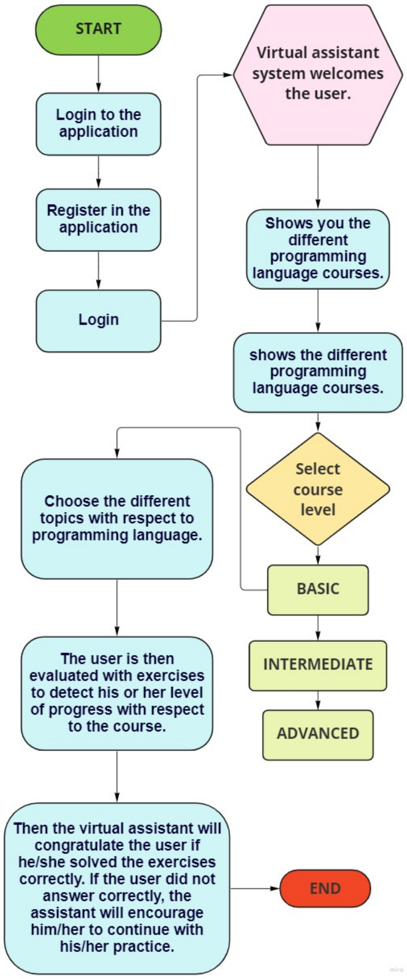

START
Login to the application
Register in the application
Login
Virtual assistant system welcomes the user.
Shows you the different programming language courses.
shows the different programming language courses.
Select course level
BASIC
INTERMEDIATE
ADVANCED
Choose the different topics with respect to programming language.
The user is then evaluated with exercises to detect his or her level of progress with respect to the course.
Then the virtual assistant will congratulate the user if he/she solved the exercises correctly. If the user did not answer correctly, the assistant will encourage him/her to continue with his/her practice.
END

Fig. 2. Flow Diagram.

Fig. 4 shows the welcome screen where the user will enter the username and password established so that the chatbot can then welcome him by greeting him with his respective name, the virtual assistant will accompany the user in his learning process so that the teaching is more didactic and the student can better grasp the different topics provided by the application.

Fig. 5 shows the different programming languages where the user can choose the course he/she wants to learn. Once the free course has been chosen, the user will be asked whether he/she wants to start at the basic, intermediate or advanced level.

Fig. 6 shows the different topics of the course where, after having studied with them, you are given exercises to put into practice the skills you have learned.

Fig. 7 shows the results of the survey where students of the University of Sciences and Humanities of the systems and computer engineering career were asked if they like programming, thus obtaining a result of 39%, they were also asked about the different difficulties in understanding the programming course, where they answered that they had problems in understanding the process, thus obtaining 31%;

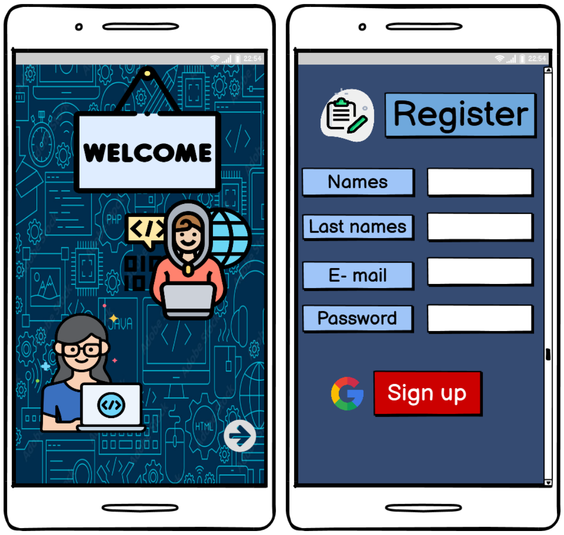
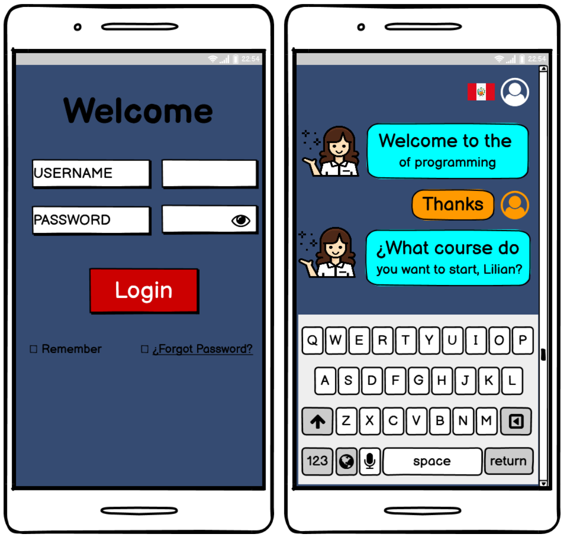


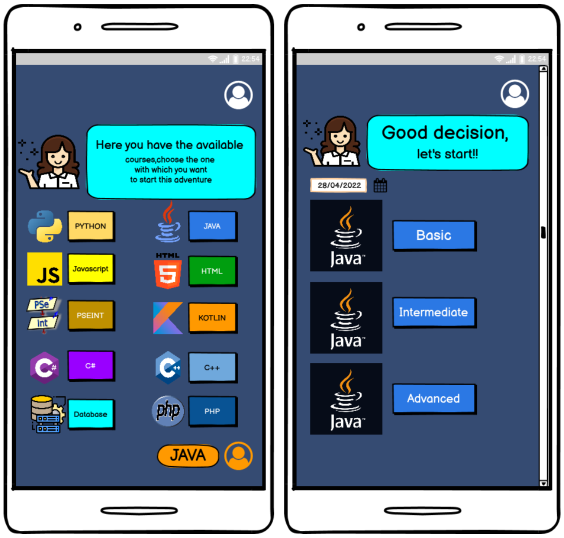
Fig. 3. Registration.



Welcome to the of programming

¿What course do you want to start, Lilian?

Fig. 4. Welcome.

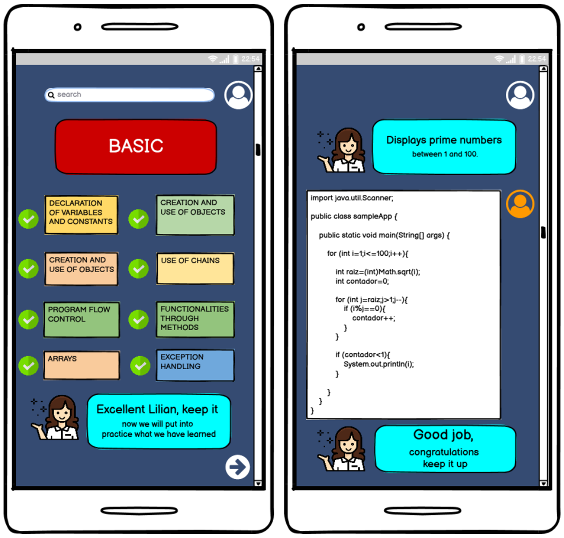


Here you have the available courses,choose the one with which you want to start this adventure

Good decision, let's start!!

28/04/2022

Fig. 5. Programming Languages.


search
BASIC
DECLARATION OF VARIABLES AND CONSTANTS
CREATION AND USE OF OBJECTS
CREATION AND USE OF OBJECTS
USE OF CHAINS
PROGRAM FLOW CONTROL
FUNCTIONALITIES THROUGH METHODS
ARRAYS
EXCEPTION HANDLING

Excellent Lilian, keep it now we will put into practice what we have learned

Displays prime numbers between 1 and 100.

```
import java.util.Scanner;

public class sampleApp {

    public static void main(String[] args) {

        for (int i=1;i<=100;i++){

            int raiz=(int)Math.sqrt(i);
            int contador=0;

            for (int j=raiz;j>1;j--){
                if (i%j==0){
                    contador++;
                }
            }

            if (contador<1){
                System.out.println(i);
            }

        }
    }
}
```

Good job, congratulations keep it up

Fig. 6. Course Level.

they were also asked if they had difficulties in programming, thus obtaining 30%.

Fig. 8 shows the results as first question 36% answering that algorithms have great relevance in the creation of something new, as second question 18% mentioning that algorithms are created by means of sequences or processes that must be executed efficiently, as third question 18% were obtained by means of tools such as the different languages that exist, also as fourth question 28% were obtained where students mention that they use Java and Python.

In Fig. 9, we return with a 35% mentioning that students use their device most of the time, also as a second question obtained a 31% where students mentioned that they consider the innovative idea of using a mobile application to help reinforce their learning based on programming and finally as a third question obtained a 34% having as a result that if it would be very useful to use the mobile application to help them to increase their learning ability related to the different programming languages.

B) About the Methodology

The Design Thinking methodology has the possibility of approaching the creations in different fields in an adaptable and extreme way. Design Thinking also has a sequence of tools that are applied throughout the process of building the innovative product, therefore it can be used constantly because it is based on the resolution of problems from the customer's point of view, that is why the Design Thinking methodology was chosen because it is considered quite suitable to propose IT resolutions in process models in which prototypes are applied to clarify the requirements to the customer's satisfaction, some of these process models are used as prototypes for their solution. They allow a great relationship with the users, such as the construction and evaluation of the elaborated prototypes.

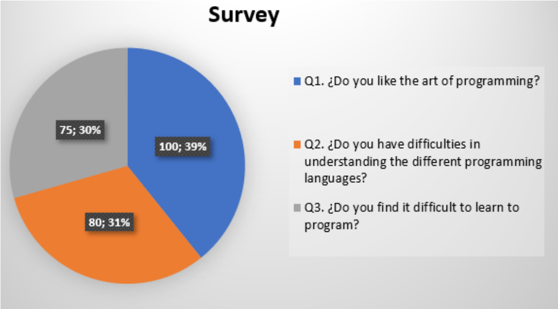


Fig. 7. Results on the Taste of Programming.

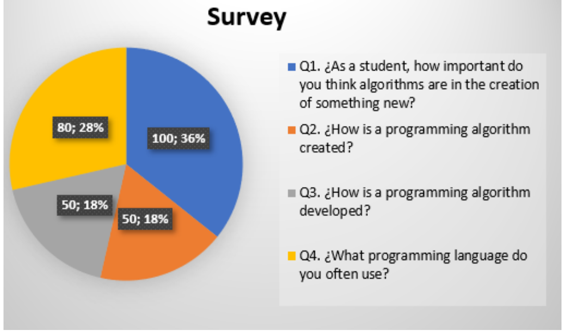


Fig. 8. Results on the Importance of Programming.

1) Advantages: One of the benefits of using this Design Thinking methodology is that it promotes the immediate solution to the problems that may arise during the development of the project. This is why the client is placed as the center of the construction process, in which the development is clear to overcome the challenges by guiding on the techniques and tools to use to solve a challenge both in an organization and in other spaces. It is necessary to emphasize that this methodology is people-centered, this means that it is a user-centered procedure whose fundamental potential is to solve real problems, also another benefit is that it adapts to the solution of the product and service through the customer's needs, thus excellent results are obtained.

2) Disadvantages: A disadvantage of the methodology is that it cannot be used for all types of projects, so it should be used in projects that require design thinking steps.

3) Comparison: Design Thinking is not only used by companies, in such a way that prestigious organizations have discovered its benefits and have incorporated it into their daily work because it designates the user as the main axis of the construction process before other approaches that try to move from thought to action. It also invites the user to be able to accept a more active role in the design of the required product, which involves the agents to dialogue between the user and the person in charge of making the innovative proposal, therefore the first approach of a team in innovation, it is essential to perform it in a focused way to a specific problem and then use the Design Thinking methodology offering a process where the stages have the possibility of being retaken and reiterated without further major limitations, In Scrum, you can work with a constant flow of projects that have to answer to the priorities through sprints, so there is a sequence of meetings and therefore there is a point where there may be a contrast, however, have the possibility to take certain resources of Scrum as prioritization, or adopt organizational points that contribute to group work.

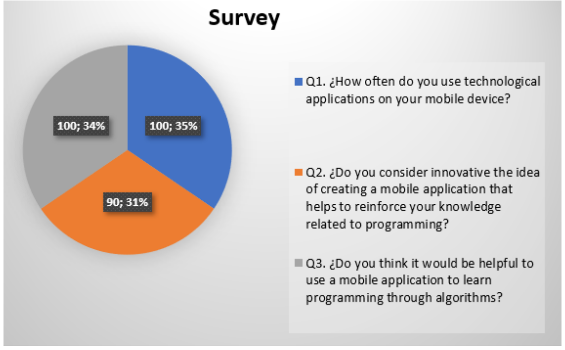


Fig. 9. Results on the Mobile Application.

TABLE IV. COMPARISON

Design Thinking	Cascade	Scrum
They focus on the users to obtain better results.	Clear structure of the methodology	It is required for large projects.
Team collaboration in such a way accelerates the cycle for the development of new solutions.	It presents difficulty in the required changes.	Provides a democratic approach.
Innovative ideas for good control and performance.	Excludes the customer or end user.	Requires a thorough homework review.
Reduces project costs and risks.	Determines the target immediately.	Requires those who use the scrum methodology, have a high level of qualification.

Table IV shows in detail the comparison of the Design Thinking, Cascade and Scrum methodologies, identifying that Design Thinking reduces risks at low cost and is adaptable to changes, and the cascade methodology presents difficulties when making required changes, and finally the scrum methodology requires a thorough review when performing tasks.

TABLE V. TRADITIONAL METHODOLOGY VS. AGILE METHODOLOGY

Traditional Methodology	Agile Methodology
The project is carried out without divisions.	The project is subdivided into different parts.
Extensive documentation is available.	There is little documentation.
Develops in a predictive manner.	It is developed in a way that is adaptable to the project
Software deliveries at the end of the project.	software deliveries are constant.
There is little communication with the client or user.	There is constant communication with the project.
Hide the error.	Immediately detects the error for resolution.

Table V shows the comparison of using an agile vs. traditional methodology, making reference that for this research the agile methodology was used since it fits the project and also has better processes to be implemented.

VI. Conclusion

In conclusion, the design of the application will help many students of the first cycles of the career of computer science and computing since they have various difficulties as shown in the survey is why it was analyzed in detail and devised the prototype of the mobile application working together with a chatbot which will help the user to perform their learning correctly in order to obtain favorable results, in addition, its teaching will be didactic which many students like the pedagogical and not the monotonous that are simple applications where they will not have motivation for learning and what is required is to help the student and facilitate their teaching complemented with the various courses of programming language for free that provides the application, on the other hand, The use of the Design Thinking methodology was satisfactory, because it focuses on the user and on the problems that may arise through the realization of the project, therefore, what made possible is the development of the prototype design of the application also obtaining data on the learning of programming in computer science students of the first cycles at the University of Sciences and Humanities. With this design of the application, it is intended in the future to have the implementation of the software to be able to implement it, in such a way to help many students or people who want to learn more about programming being in different countries benefiting their knowledge and being enriched every day more through the knowledge provided.

VII. Future Work

With this design of the application, it is intended in the future to have the implementation of the software to be able to implement it, in such a way to help many students or people who want to learn more about programming being in different countries benefiting their knowledge and being enriched every day more through the knowledge provided. With this design of the application, it is intended in the future to have the implementation of the software to be able to implement it, in such a way to help many students or people who want to learn more about programming being in different countries, benefiting their knowledge and getting richer every day through of the knowledge provided. In addition, it must be applied to other courses of complexity in learning; for this they must select a complexity course for their application.

Acknowledgment

This research work was supported by the University of Sciences and Humanities with its respective research institute.

References

[1] A. Gruenewald, C. Giesser, S. Buechner, C. Gibas, and R. Brueck, "Going virtual: Teaching practical skills of circuit design and programming for heterogeneous groups online," in *2021 IEEE Global Engineering Education Conference (EDUCON)*, 2021, pp. 404–412. [Online]. Available: https://doi.org/10.1109/EDUCON46332.2021.9454125

[2] A. D. Rio-Chillcce, L. Jara-Monge, and L. Andrade-Arenas, "Analysis of the use of videoconferencing in the learning process during the pandemic at a university in lima," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, pp. 870–878, 2021, doi:10.14569/IJACSA.2021.01205102.

[3] A. R. Bernaola, M. A. Tipula, J. E. Moltalvo, V. S. Sandoval, and L. Andrade-Arenas, "Analysis of the use of technological tools in university higher education using the soft systems methodology," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, pp. 412–420, 2020, doi:10.14569/IJACSA.2020.0110754.

[4] H. Amer and S. Harous, "Smart-learning course transformation for an introductory programming course," in *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 2017, pp. 463–465. [Online]. Available: https://doi.org/10.110/ICALT.2017.91

[5] L. Andrade-Arenas, D. L. Nunez, and C. Sotomayor-Beltran, "Leveraging digital tools for a better virtual teaching-learning process in a private university of lima," in *EDUNINE 2021 - 5th IEEE World Engineering Education Conference: The Future of Engineering Education: Current Challenges and Opportunities, Proceedings*, 2021, doi:10.1109/EDUNINE51952.2021.9429113.

[6] M. B. Garcia, "Cooperative learning in computer programming: A quasi-experimental evaluation of jigsaw teaching strategy with novice programmers," *Education and Information Technologies*, vol. 26, no. 4, pp. 4839–4856, 2021. [Online]. Available: https://doi.org/10.1007/s10639-021-10502-6

[7] J. S. Amro and R. Romli, "Investigation on the learning programming techniques via mobile learning application," in *2019 4th International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*, 2019, pp. 1–7. [Online]. Available: https://doi.org/10.1109/ICRAIE47735.2019.9037764

[8] N. F. Rozali and N. M. Zaid, "Code puzzle: Actionscript 2.0 learning application based on problem based learning approach," in *2017 6th ICT International Student Project Conference (ICT-ISPC)*, 2017, pp. 1–4. [Online]. Available: https://doi.org/10.1109/ICT-ISPC.2017.8075329

[9] P. H. Coronel, T. Alejandra Loor Rengifo, M. A. Henríquez Coronel, J. Pablo Trampuz Reyes, and I. F. Fernández, "Information literacy in latin america students: a review of programs and proposes," in *2020 XV Conferencia Latinoamericana de Tecnologias de Aprendizaje (LACLO)*, 2020, pp. 1–7. [Online]. Available: https://doi.org/10.1109/LACLO50806.2020.9381144

[10] C. Sotomayor-Beltran, G. W. Z. Segura, and A. Roman-Gonzalez, "Why should python be a compulsory introductory programming course in lima (peru) universities?" in *2018 IEEE International Conference on Automation/XXIII Congress of the Chilean Association of Automatic Control (ICA-ACCA)*, 2018, pp. 1–4. [Online]. Available: https://doi.org/10.1109/ICA-ACCA.2018.8609808

[11] R. P. Curasma, K. O. Villalba-Condori, N. J. Jara, R. Q. Llamoca, J. C. C. Chávez, and M. D. P. Ponce-Aranibar, "Computational thinking and block-based programming for beginning engineering students: Systematic review of the literature," in *2021 XVI Latin American Conference on Learning Technologies (LACLO)*. IEEE, 2021, doi: 10.1109/LACLO54177.2021.00096, pp. 530–533.

[12] L. Andrade-Arenas and C. Sotomayor-Beltran, "On the perspectives of graduated engineering students on three dimensions of the integrated curriculum from a peruvian university," in *Proceedings of the 2019 International Symposium on Engineering Accreditation and Education, ICACIT*, 2019, doi:10.1109/ICACIT46824.2019.9130268.

[13] J. Sitompul, "Student perceptions of the use of android-based learning media in the production ecrite intermediaire course," *Budapest International Research and Critics in Linguistics and Education (BirLE) Journal*, vol. 3, 2020. [Online]. Available: https://doi.org/10.33258/birle.v3i1.859

[14] B. F. Springgate, A. C. Arevian, A. Wennerstrom, A. J. Johnson, D. P. Eisenman, O. K. Sugarman, C. G. Haywood, E. J. Trapido, C. D. Sherbourne, A. Everett, M. McCreary, D. Meyers, S. Kataoka, L. Tang, J. Sato, and K. B. Wells, "Community resilience learning collaborative and research network (c-learn): Study protocol with participatory planning for a randomized, comparative effectiveness trial," *International Journal of Environmental Research and Public Health*, vol. 15, 2018. [Online]. Available: https://doi.org/10.3390/ijerph15081683

[15] N. Parveen and S. Zamir, "Factors affecting behavioural intentions in the use of mobile learning in higher education," *International Journal of Distance Education and E-Learning*, vol. 6, 2021. [Online]. Available: https://doi.org/10.36261/ijdeel.v6i1.1430

[16] S. Mahadevappa and S. Figueira, "Energy-efficient programming languages for mobile applications," in *2021 IEEE Global Humanitarian Technology Conference (GHTC)*, 2021, pp. 33–38. [Online]. Available: https://doi.org/10.1109/GHTC53159.2021.9612479

[17] J. C. Paiva, J. P. Leal, and A. Figueira, "Automated assessment in computer science education: A state-of-the-art review," *ACM Trans. Comput. Educ.*, jan 2022, just Accepted. [Online]. Available: https://doi.org/10.1145/3513140

[18] L. G. Martínez, S. Marrufo, G. Licea, J. Reyes-Juárez, and L. Aguilar, "Using a mobile platform for teaching and learning

object oriented programming," *IEEE Latin America Transactions*, vol. 16, no. 6, pp. 1825–1830, 2018. [Online]. Available: https://doi.org/10.1109/TLA.2018.8444405

[19] Y. Ouyang, B. Guo, Q. Wang, Y. Liang, and Z. Yu, "Learning dynamic app usage graph for next mobile app recommendation," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2022. [Online]. Available: https://doi.org/10.1109/TMC.2022.3161114

[20] E. Bone, D. Evitaputri, and P. Santaanop, "Mobile learning in higher education environmental science: state of the field and future possibilities," *Pacific Journal of Technology Enhanced Learning*, vol. 4, no. 1, pp. 1–3, Jan. 2022, doi: 10.24135/pjtel.v4i1.123.

[21] A. Yassine, M. Berrada, A. Tahiri, and D. Chenouni, "A cross-platform mobile application for learning programming basics," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 12, no. 7, p. pp. 139–151, Nov. 2018. [Online]. Available: https://online-journals.org/index.php/i-jim/article/view/9442

[22] D. M. Alghazzawi, S. H. Hasan, G. Aldabbagh, M. Alhaddad, A. Malibari, M. Z. Asghar, and H. Aljuaid, "Development of platform independent mobile learning tool in saudi universities," 2021. [Online]. Available: https://doi.org/10.3390/su13105691

[23] S. Siyu, "A application study of artificial intelligence aided design," in *2020 International Conference on Innovation Design and Digital Technology (ICIDDT)*, 2020, pp. 98–101. [Online]. Available: https://doi.org/10.1109/ICIDDT52279.2020.00026

[24] X. He, L. Xu, X. Zhang, R. Hao, Y. Feng, and B. Xu, "Pyart: Python api recommendation in real-time," in *2021 IEEE/ACM 43rd International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, 2021, pp. 246–247. [Online]. Available: https://doi.org/10.1109/ICSE-Companion52605.2021.00114

[25] J. Park, J. Park, S. An, and S. Ryu, "Jiset: Javascript ir-based semantics extraction toolchain," in *Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering*, ser. ASE '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 647–658. [Online]. Available: https://doi.org/10.1145/3324884.3416632

[26] M. Sãnchez, E. V. Bahamondez, and G. T. de Clunie, "Use of pseint in teaching programming: A case study," in *Proceedings of the 10th Euro-American Conference on Telematics and Information Systems*, ser. EATIS '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3401895.3402083

[27] E. Buonanno, *Functional Programming in C*. Simon and Schuster, 2022.

[28] W. Yandong, "Research on the fragmented learning of "java language programming" in the internet + era," in *2021 16th International Conference on Computer Science Education (ICCSE)*, 2021, pp. 375–378. [Online]. Available: https://doi.org/10.1109/ICCSE51940.2021.9569637

[29] P. Skibinski, "Improving html compression," in *Data Compression Conference (dcc 2008)*, 2008, pp. 545–545. [Online]. Available: https://doi.org/10.1109/DCC.2008.74

[30] S. M. Lucas, "Cross-platform games in kotlin," in *2020 IEEE Conference on Games (CoG)*, 2020, pp. 774–775. [Online]. Available: https://doi.org/CoG47356.2020.9231914

[31] D. Powers, *What Is PHP 8?* Berkeley, CA: Apress, 2022, pp. 1–7. [Online]. Available: https://doi.org/10.1007/978-1-4842-7141-4-1

Prediction of COVID-19 Patients Recovery using Ensemble Machine Learning and Vital Signs Data Collected by Novel Wearable Device

Hasan K. Naji[1]
School of Automatic Control and Computers
University POLITEHNICA of Bucharest
Bucharest, Romania

Hayder K. Fatlawi[2]
Department of Information Systems
ELTE Eötvös Loránd University
Budapest 1117, Hungary
IT Research and Development Centre
University of Kufa
Najaf, Iraq

Ammar J. M. Karkar[3]
IT Research and Development Centre
Electronic and Communication Department
University of Kufa
Najaf, Iraq

Nicolae GOGA[4]
Faculty of Engineering in Foreign Languages
University POLITEHNICA of Bucharest
Bucharest, Romania
Molecular Dynamics Group
University of Groningen
Groningen, Netherlands

Attila Kiss[5]
Department of Information Systems
ELTE Eötvös Loránd University
Budapest 1117, Hungary
Department of Informatics
J. Selye University,
Komárno 94501, Slovakia

Abdullah T. Al-Rawi[6]
Al-Karkh General Hospital
Baghdad, Iraq

Abstract—**During the spread of a pandemic such as COVID-19, the effort required of health institutions increases dramatically. Generally, Health systems' response and efficiency depend on monitoring vital signs such as blood oxygen level, heartbeat, and body temperature. At the same time, remote health monitoring and wearable health technologies have revolutionized the concept of effective healthcare provision from a distance. However, analyzing such a large amount of medical data in time to provide the decision-makers with necessary health procedures is still a challenge. In this research, a wearable device and monitoring system are developed to collect real data from more than 400 COVID-19 patients. Based on this data, three classifiers are implemented using two ensemble classification techniques (Adaptive Boosting and Adaptive Random Forest). The analysis of collected data showed a remarkable relationship between the patient's age and chronic disease on the one hand and the speed of recovery on the other. The experimental results indicate a highly accurate performance for Adaptive Boosting classifiers, reaching 99%, while the Adaptive Random Forest got a 91% accuracy metric.**

Keywords—Machine learning; COVID-19; wearable device

I. Introduction

Pandemics produced by infectious diseases always had a negative impact on the community at large. Currently, the globe is witnessing the advent of SARS linked to a novel coronavirus (SARSCov2) [1]. Patients with suspected exposure or symptoms should be identified as soon as possible. In respiratory infections, vital signs such as body temperature, heartbeat, and blood oxygen saturation are acknowledged as crucial strong indicators [2], [3]. Their monitoring is crucial for detecting early physiological abnormalities in patients who are deteriorating. They play a critical role in triaging patients for proper care and predicting whether they will improve or deteriorate [4], [5].

Approximately 82% of SARSCov2, also known as COVID-19, patients have minimal symptoms, recover quickly, and do not need to be hospitalized. 10 to 20% of patients who require hospitalization require care in intensive care units (ICUs), 3 to 10% require intubation, and the case fatality rate ranges from 2 to 5% [6]. Patients with mild illness are typically treated symptomatically, with home isolation as the last resort [7]. COVID-19's incubation period (the time between infection and beginning of symptoms) ranges from 2 to 14 days, with an average of (5 to 6) days [8]. Vital signs have to be monitored by the patients who have been assigned to a quarantine zone.

Wearable remote patient monitoring systems have made it possible to monitor vital indicators in a regular basis outside medical facilities and/or where mitigating contact with health workers is required [9]. However, with long-term monitoring and vital signs recording, a large amount of data is generated continuously, and health workers find it hard to extract information. This information is crucial to setting health policies and determining patient treatments. Thus advancements in machine learning combined with medical systems have resulted in a revolutionary discovery [5]. Commercial introductions of such gadgets without medical validation studies have occurred recently, rendering them inappropriate for medical use [10].

Generally, the classification process in data mining aims for the descriptive or predictive task. A descriptive analysis may be utilized to illustrate data relationships in a way that decision-makers can comprehend. In predictive analysis, the classification model can also be used for forecasting future values for the target class. One of the most popular data mining techniques is the decision tree DT. It is a simple and effective method that is used for both tasks. In [11], the proposed system presents a wearable device capable of monitoring the patient's vital signs (body temperature, oxygen saturation, heartbeat) and analyzing these data to predict the patient's recovery using ensemble classifiers with a DT as a based learner. The main contributions of this manuscript is:

- Utilizing an IoT-based monitoring health system to collect COVID-19 patients data set.
- Analysing the collected data and extracting crucial medical information using ensemble machine learning techniques.
- Evaluating and discussing the key findings of the main features of the medical data set.

The rest of this research is organized as follows: Section II reviews some related works. Section III describes the proposed classification system architecture along with the deigned wearable device. Section IV discuss the collected patient dataset and its characteristics, while Section V analyze this dataset and the performance of ensemble classifiers. Section VI discuss the results, limitations, and possible applications of the proposed system. Lastly, Section VII summarizes the main conclusions of this work.

II. Related Work

A. Health Monitoring Systems

Recent advancements in IoT and wearable device technologies enabled significant improvements in health applications, especially in the current pandemic that has become a global issue and threat to public health. For instance, Albassam $etal.$ Proposes an IoT-based health monitoring system for COVID-19 patients to measure various vital signs such as temperature, heart rate, oxygen saturation, and cough count, as well as to report patient GPS location data to medical authorities in real-time [12]. The system includes a wearable body sensor, web API, and a mobile application. The monitoring system is connected to the (IoT) cloud, where data is processed and analyzed.

Moreover, A real-time wearable monitoring system for the COVID-19 patient had been proposed [13]. A wearable chest patch and a pulse oximeter are used to send patients vital signals, including heart rate, respiration rate, and peripheral oxygen saturation, via wireless Android tablet devices to the nursing ward. The proposed system offers a technical wearable device (bracelet) to make it more unrestricted and give patients complete freedom of movement and do all activities without restrictions inside the home or hospital. The system added the technology of quarantine monitoring for those infected with the virus by adding GSM technology.

On the other hand, Mizher $et\,al.$ propose an internet of medical things (IoMT) based healthcare monitoring system that uses a wearable device [14]. This wearable device includes two sensors to measure blood oxygenation, temperature, and heart rate to monitor the health status of COVID-19 patient and limits virus spread. Their proposed system provides more services at a lower cost with the ability to change functions according to the requirements of medical staff, whether during the epidemic period or afterward.

Gloria.C $etal.$ [15] proposes a systematic review on accuracy and metrological characteristics of wrist-worn and chest-strap wearable devices. According to this system, aspects such as calibration procedure, number of test protocols, measured quantities, absolute error percentage, and correlation coefficient should be considered to evaluate the accuracy of wearable devices. The evaluation of the accuracy of wearable devices performed by a methodology was based on the calibration technique, the number of test protocols, measured quantities, absolute error rate, and correlation coefficient. However, in all the mentioned studies, either the wearable device design requires further efficiency or the produced health data lack a smart data analysis for this nontrivial big stream of data. This study aim to address this by proposing an efficient and smart health system.

B. Machine Learning in Health Applications

In order to extract or predict useful information from medical health dataset, many studies had proposed machine learning techniques. For instance, two machine learning techniques had been utilized for classifying COVID-19 against influenza data [16]. The first stage of their model was to make two clusters using Fuzzy C-Mean, then a Back Propagation classifier was built based on the clustering result. The implementation of their model included creating a mobile application that receives medical test data from the user and applies the classification within the mobile environment. Another classification framework proposed by [17] for detecting COVID-19 infection in the chest radiograph. X-ray images are classified using a convolutional neural network (CNN) and generative adversarial network (GAN) for data augmentation in their framework.

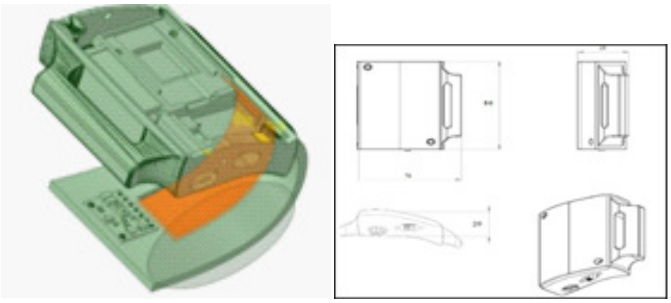
A smartphone-based recognition classifier is proposed by [18] for differentiating the normal coughs and uninfected persons from COVID-19 infected ones. Their work included a comparison of seven classification techniques, and the results showed that the neural network based on residual had the best performance. The author in [19] proposed a CNN classification model for classifying Computer Tomography (CT) images of COVID-19 patients. Their model required a low computational resource as it can be implemented in a personal computer without GPU acceleration. However, the works mentioned in this section are concerned with detecting COVID-19 infection, not the recovery status. In addition, two of them used medical images that are only available in the medical centers.

III. Proposed System Architecture

The proposed health classification system continuously records the sensors' output values alongside other medical information for COVID-19 patients and predicts their recovery. In addition, the system monitors the status of a patient; whenever the system recognizes a predefined critical level for any of the patient's vital signs, the system will send an alert to the responsible health staff. This system contains various components: (1) the Bracelet, which contains sensors for reading the patient's vital signs and passing them to the basic control unit, and (2) The Base Control Unit for receiving, analyzing, classifying, storing, and displaying data for each patient, (3) the user interface is designed to arrange and engage all system functions, inputs, and outputs for patients and users [9]. Moreover, it displays the instant patient vital signs to health workers.

A. Bracelet Design

Wearable electronics devices are starting to include many fundamental and even leisure functionalities due to market demands. In this research, the wearable device, which is the bracelet, has been designed to perform a vital task in the patient monitoring system. When constructing such systems, stability, durability, safety, and acceptable form factors must all be taken into account. Several layouts were implemented in this initial investigation, and the one illustrated in Fig. 1 was selected [20].



(a) (b)

Fig. 1. Bracelet Designed Layout.

The microprocessor, battery, sensors, and other components are all housed in the bracelet case, which measures (86 x 70 x 29) mm. Slots are cut into the lower and top parts of the bracelet so that the sensors can be placed on the skin with the needed accuracy, as shown in Fig. 1, the bracelet (container) is expertly created for patients of all ages and genders. The bracelet's exterior had an efficient technical curve that was made with a 3D printer with flexible materials to match the size and shape of most users' wrists. In addition, this design has shown the best accuracy in recording vital signs and communicating with the base station [20].

B. Building Ensemble Classifier

The data of a classification task with the size N consists of a set of features X $(x1, ..., xM)$ and Y, which is the class label vector, where N is the number of instances and M is the number of features. This task aims to build a model that can classify or predict the value of y depending on the given feature set X. The single model can produce unreliable predictions because of overfitting, so the ensemble classification includes building multiple weak classifiers (base learners) and merging their decisions to make one strong one. In the proposed system, the base learner is the decision tree consisting of two types of nodes; the internal node represents a condition on a data feature used to divide the data records. Only two branches produce from each internal node based on the condition in a binary tree, while the multi-branch DT can generate more than two branches from each internal node. A leaf node is a non-internal that holds a class label without any data splitting.

According to the mechanism of merging the base learners, ensemble classification is categorized as Boosting and Bagging, and both of them are used in the proposed system. AdaBoost refers to Adaptive Boosting [21]; it performs by assigning a higher weight to data records that are incorrectly classified and less weight to those that are already well-classified. This process is performed by many iterations, which is called the size of the ensemble, and it is equal to the number of base learners. For the initial iteration, the data subset is chosen randomly with the equal possibility of choosing each instance. In the following iterations, the weight of an instance will be increased if it was incorrectly classified in the previous iteration. AdaBoost predicts the unknown value for any new instance by using the weighted average of base learners' decisions. The weight of each learner is calculated base on its misclassification error. Fig. 2 illustrated steps of building the proposed ensemble classifier based on AdaBoost algorithm.

Random forest applies Bagging in which a random subsampling is chosen with a replacement in each iteration for the training set. In addition, a random subset of features will be selected in the splitting process of DT. Adaptive random forest ARF [22] is dedicated to dealing with data streams by applying a continual training method for the classification model to adjust to changes in data distribution. It uses a modified version of DT called Hoefding Tree of Very Fast Decision Tree, which can adapt itself as a response to changes in data distribution. Adaptive Window ADWIN is used in ARF for monitoring and detecting that change in a data stream.

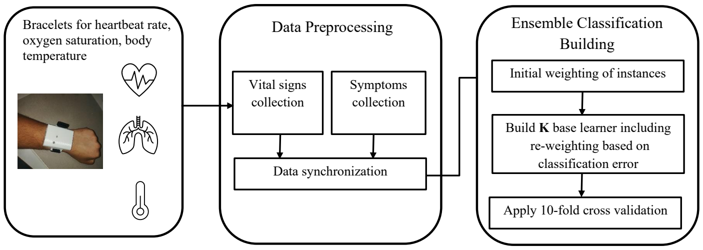

Bracelets for heartbeat rate, oxygen saturation, body temperature
Data Preprocessing
Vital signs collection
Symptoms collection
Data synchronization
Ensemble Classification Building
Initial weighting of instances
Build **K** base learner including re-weighting based on classification error
Apply 10-fold cross validation

Fig. 2. Steps of Building the Proposed Ensemble Classifier

IV. Patients Data Sample

The study sample in this research is made up of a voluntary group of individuals infected with COVID-19 found inside clinics designated to treat patients with COVID-19. In addition to the vital signs from the bracelet, the patient's medical history, how long they have been infected, and their symptoms are collected with the help of health workers in these clinics. Thus, each patient record is designed to show the pathological behavior of the virus throughout the infection period. Therefore, using data analysis techniques, information might be extracted to prevent patients from reaching critical conditions.

A. Sample Demography

The system was tested on a non-governmental voluntary sample targeting patients infected with the virus during the pandemic, which had mild or moderate symptoms while attending outpatient clinics in Iraq, in the capital, Baghdad. The sample included (408) infected individuals, and their information and medical history were recorded without mentioning their names in order to preserve privacy. The collected data has been divided into two datasets; the first one, entitled Covid19-IQ01, contains data from 313 patients, one record for each patient. The second dataset, Covid19-IQ02 created by monitoring 95 patients for five days, i.e., five records for each patient and 475 in total. In the Covid19-IQ01 data set, gender is fairly distributed, as shown in Table I. In addition, the sample data is categorized into six age groups, and since most of them suffered from high to moderate infection, the age group (60-69 years old) is the highest with (27.5%). Table I also shows that 55% of 313 patients have some kind of chronic disease, and this can provide the ability to indicate their impact on recovery period of COVID-19 patients.

TABLE I. The Demography Characteristics of Sample Patients

		Covid19-IQ01	**Covid19-IQ02**
Gender	Male	≈ 53%	≈ 45.2%
	Female	≈ 47%	≈ 44.8%
Age	10-19	≈ 16.6%	≈ 17.9%
	20-29	≈ 15.6%	≈ 12.6%
	30-39	≈ 14%	≈ 16.8%
	40-49	≈ 13.4%	≈ 16.8%
	50-59	≈ 12.8%	≈ 17.9%
	60-69	≈ 27.5%	≈ 17.9%
Chronic diseases	Hypertension	≈ 28.4%	≈ 27.3%
	Diabetes	≈ 28.4%	≈ 30.5%
	Others	≈ 8%	≈ 12.6%
	None	≈ 45.3%	≈ 46.3%
Symptoms	Exhaustion	≈ 55.5%	≈ 62%
	Dizziness	≈ 61%	≈ 65%
	Fast Heartbeat	≈ 59.1%	≈ 65%
	Depression	≈ 35.4%	≈ 69%

In addition, the system was tested on a sample of (95) volunteer participants, each of whom had (5) serial measurements starting from the fifth to the tenth day and had (mild or moderate) symptoms of infections. This is due to the fact that this period represents the tipping point for various complications, especially for older patients and those with chronic diseases, such as high blood pressure, diabetes, or other [23]. This sample is divided into (six age groups) from (10-70) years and two categories based on gender (43) males and (52) females. In addition, the medical history, and chronic diseases, of the patient were adopted, as the number of participants in the sample who had high blood pressure (26) participants, diabetes (29) participants, and other diseases (12) participants. Moreover, Health experts have classified the patient's status based on interviewing the patient into three categories: Recovered, recovering, and No sign of recovery. This classification is based on the persistence of symptoms associated with the disease due to infection. Taking into account the readings of vital signs, which are summarized in Table II, especially (body temperature and the oxygen saturation in the blood). These patient data records will be used in machine learning in the next sections.

V. Experiments Results

A. Data Analysis by Visualization

TABLE II. Summery of Numerical Features of Covid19-IQ01

Attribute	***Min.***	***Max.***	***Mean***	***STD***
Age	13	69	37.42	18.21
Body Temperature	36.2	38.4	37.29	0.62
Oxygen Saturation	90	97	94.62	1.5
Heartbeat Rate	55	86	69.62	6,17

In this section, we explore the patterns in the collected data Covid19-IQ01 and the relationship between the patient's condition on the one hand and age, chronic diseases, and vital signs on the other. In Fig. 3, 4, 5, 6, 7, 8, and 9,the color of the circle indicates the status of the patient, while the size of it relates incrementally to the age of the patient. RapidMiner platform [24] was used for analyzing and visualizing operations. Fig. 3 illustrates the relation between two common chronic diseases, Hypertension and Diabetes, with the progress of patient recovery. The figure showed clearly that patients with those two diseases have no sign of recovery, while there is a high ratio of recovered or recovering patients without Hypertension and Diabetes.

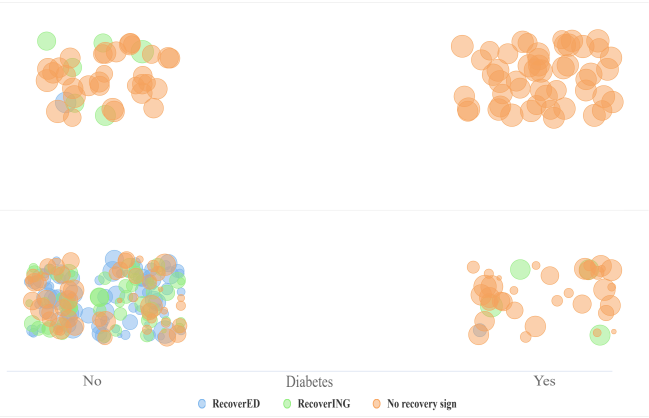


Fig. 3. Covid19-IQ01 Data Visualization: Hypertension and Diabetes

The impact of body temperature collected from the thermometer sensor on the recovery status is illustrated in Fig. 4 and Fig.5. Young Patients without Hypertension or Diabetes and 37 degrees or less of temperature mostly healed. In the case of having one of the two chronic diseases, most patients don't show a sign of recovery, although their body temperature is normal.

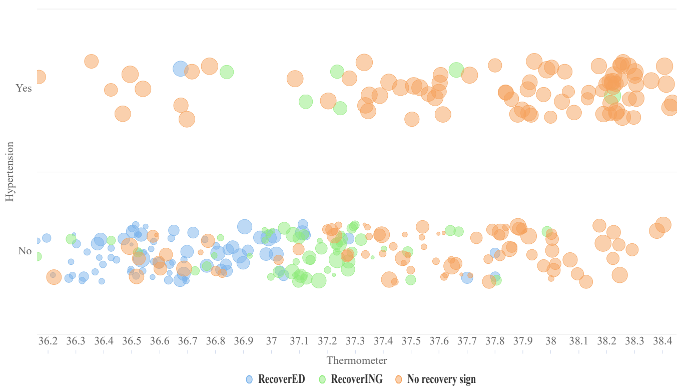


Fig. 4. Covid19-IQ01 Data Visualization: Hypertension and Thermometer.

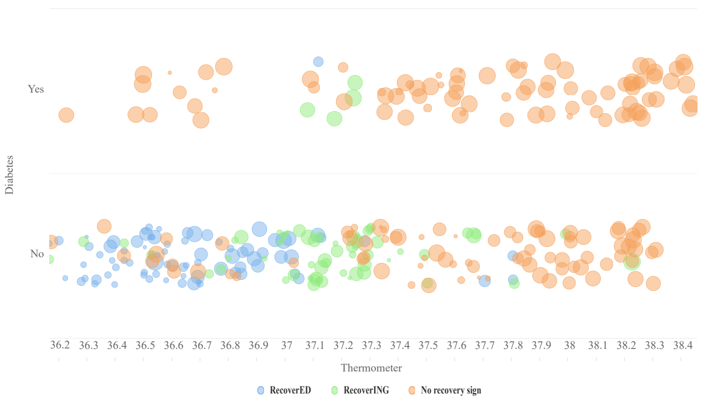


Fig. 5. Covid19-IQ01 Data Visualization: Diabetes and Thermometer.

Diabetes patients mostly have no sign of recovery regardless the age and their heartbeat rate, as shown in Fig. 6. Also, it indicates that Heartbeat rate is related to the recovery status of young patients without Diabetes. Adult patients with Hypertension mostly were not healing regardless of heartbeat rate, as shown in Fig. 7. In the normal range of heartbeat, most of the young patients were recovering or they already recovered. The same conclusions can be seen in Fig. 8 and Fig. 9 about the relation between the oxygen saturation, chronic diseases, and recovery from COVID-19.

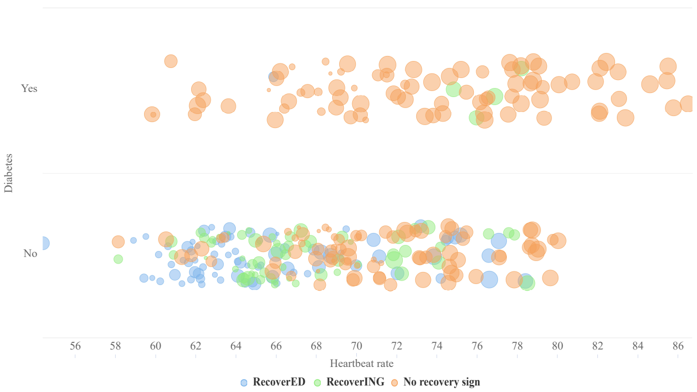


Fig. 6. Covid19-IQ01 Data Visualization: Diabetes and Heartbeat Rate.

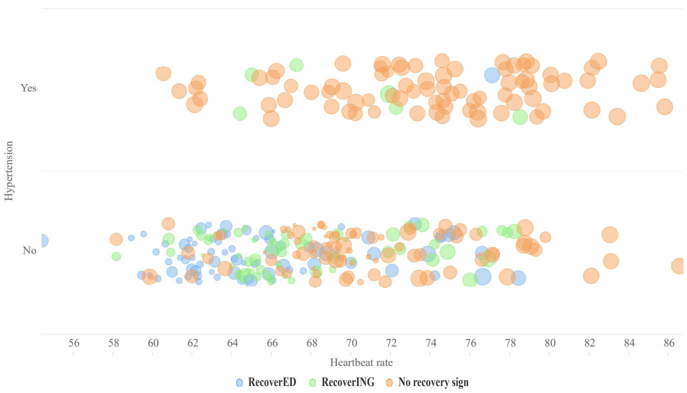


Fig. 7. Covid19-IQ01 Data Visualization: Hypertension and Heartbeat Rate.

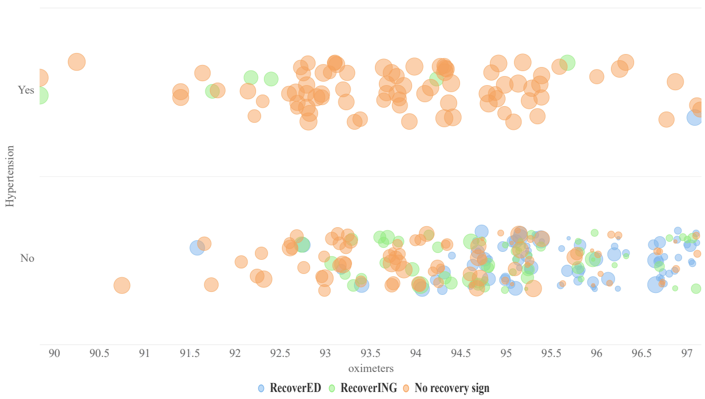


Fig. 8. Covid19-IQ01 Data Visualization: Hypertension and Oximeters.

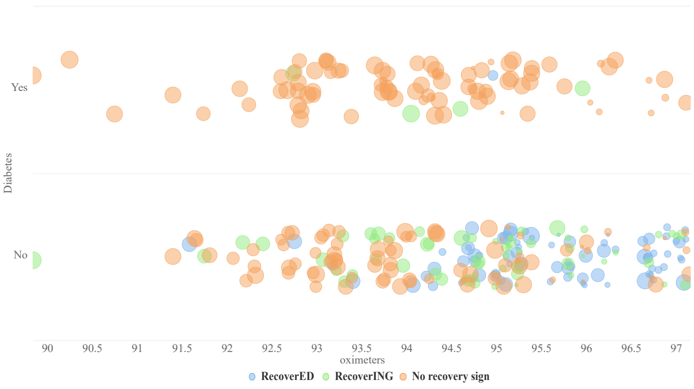


Fig. 9. Covid19-IQ01 Data Visualization: Diabetes and Oximeters.

B. Static Ensemble Classification

Waikato Environment for Knowledge Analysis [25] platform is used in this implementation for randomization, feature ranking, and classification. Implementing AdaBoost Ensemble classifier on both Covid19-IQ01 and Covid19-IQ02 datasets led to high accurate classification. The base learner for Adaboost was a J48 decision tree classifier that illustrated in Fig. 10 and Fig. 11. All the experiments in this section included 10-fold cross-validation to prevent bias, and the ensemble size was 20 for Covid19-IQ01 and 12 for Covid19-IQ02.

Tables III and IV represent the confusion matrix of AdaBoost with the two datasets; they show the correct predicted values in the bold numbers in the diagonal cells. The results in those tables indicate the accuracy of AdaBoost, especially in Table IV that contains only one record classified incorrectly.

TABLE III. CONFUSION MATRIX OF ADABOOST CLASSIFIER FOR COVID19-IQ01

	Predicted values		
Actual values	***Recovered***	***Recovering***	***No recovery sign***
Recovered	**72**	4	6
Recovering	3	**51**	9
No recovery sign	6	12	**150**

TABLE IV. CONFUSION MATRIX OF ADABOOST CLASSIFIER COVID19-IQ02

	Predicted values		
Actual values	***Recovered***	***Recovering***	***No recovery sign***
Recovered	**145**	0	0
Recovering	0	**120**	0
No recovery sign	1	0	**189**

A performance comparison is performed between AdaBoost and the other four popular classifiers. Three of them were a single classifier (J48, REPTree, and Hoeffding Tree), while the fourth is Random Forest which is an Ensemble classifier that uses bagging instead of boosting. The results of the comparison in Tables V and VI showed that Adaboost has the best results in both datasets compared with the other four classifiers based on five evaluation metrics (TP rate, FP rate, Precision, Recall, and F-Measure). Fig. 12 illustrates the importance of each feature in both datasets during the building of the AdaBoost classifier. It can be seen that AdaBoost chose a different subset from features in every dataset; only the Age feature had similar importance in both of them.

TABLE V. CLASSIFICATION PERFORMANCE OF FIVE CLASSIFIERS OF COVID19-IQ01

Technique	***TP Rate***	***FP Rate***	***Precision***	***Recall***	***F-Measure***
J48	0.837	0.103	0.838	0.837	0.838
REPTree	0.831	0.119	0.830	0.831	0.830
HoeffdingTree	0.655	0.198	0.644	0.6551	0.648
Random Forest	0.866	0.084	0.866	0.866	0.866
AdaBoost	**0.872**	**0.079**	**0.874**	**0.872**	**0.873**

TABLE VI. CLASSIFICATION PERFORMANCE OF FIVE CLASSIFIERS OF COVID19-IQ02

Technique	***TP Rate***	***FP Rate***	***Precision***	***Recall***	***F-Measure***
J48	0.977	0.010	0.977	0.977	0.977
REPTree	0.977	0.007	0.85	0.985	0.985
Hoeffding Tree	0.937	0.029	0.938	0.937	0.937
Random Forest	0.973	0.013	0.973	0.973	0.973
AdaBoost	**0.99**	**0.001**	**0.998**	**0.998**	**0.998**

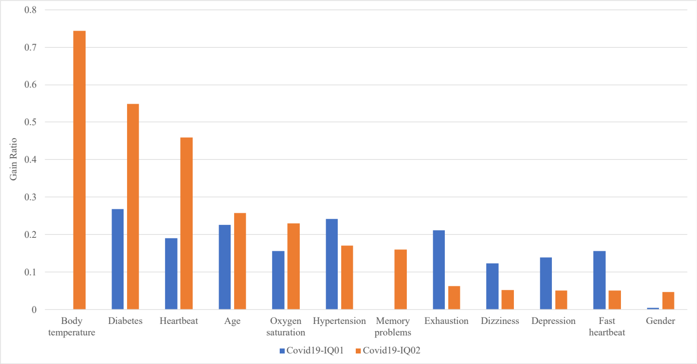


Fig. 12. Features Ranking of AdaBoost Classifer based on Gain Ratio.

C. Adaptive Ensemble Classification

Implementation of the second dataset, COVID10-IQ02, as a continuous data stream was performed in Massive Online Analysis platform and scikit-multiflow library in Python. Adaptive Random Forest ARF classifier, which is an adaptive ensemble classifier, had an incremental classification performance while receiving the stream samples with a size of 20 instances, as shown in Fig. 13. ARF obtained the best result compared with the other four adaptive classifiers by using ten base learners as shown in Fig. 14.

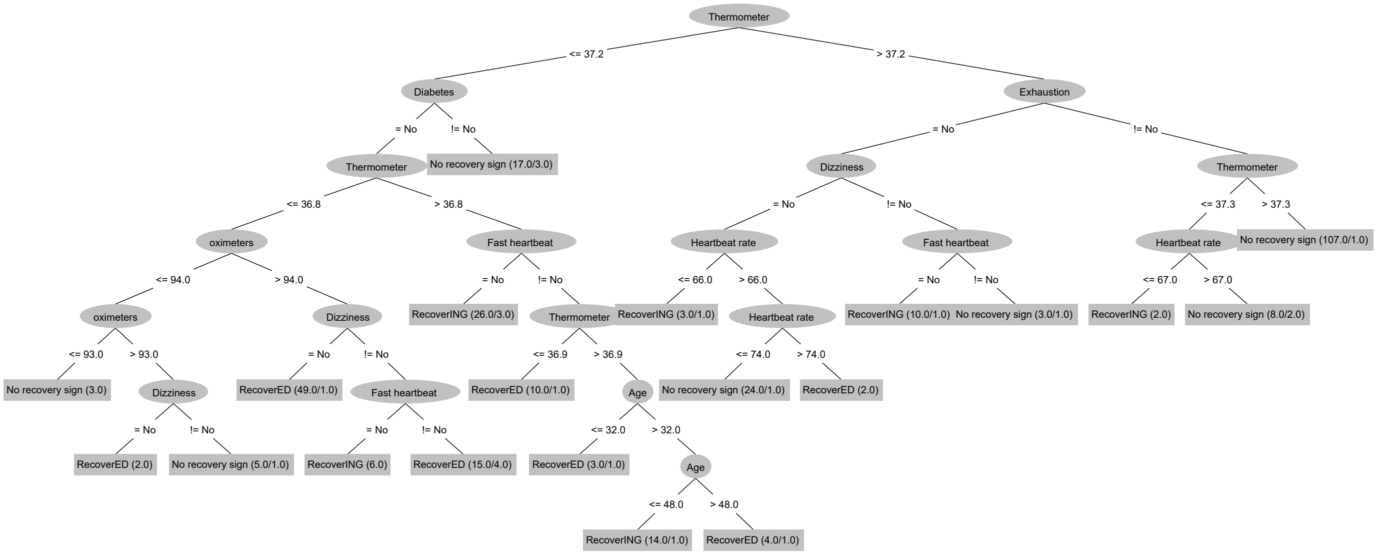


Fig. 10. Classification Tree of Covid19-IQ01 using J48 Algorithm.

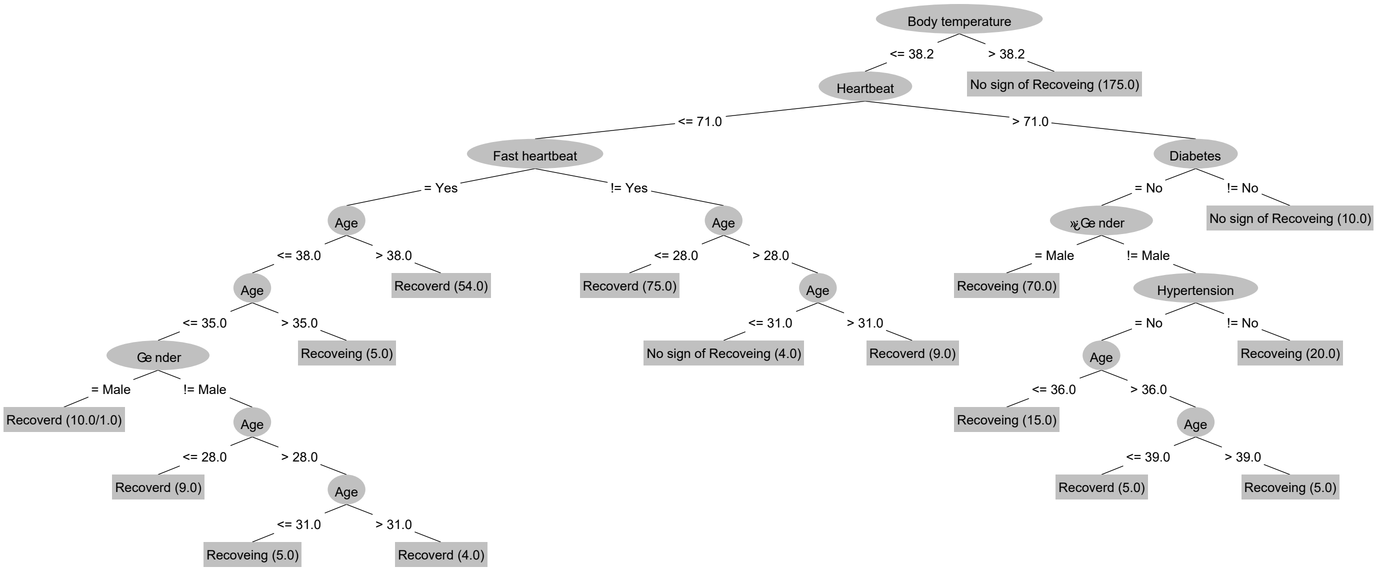


Fig. 11. Classification Tree of Covid19-IQ02 using J48 Algorithm.

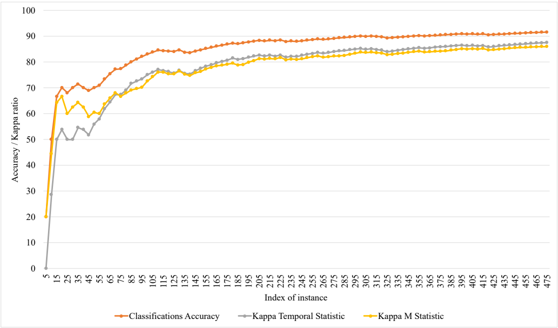


Fig. 13. Classification Performance of Adaptive Random Forest using Covid19-IQ02.

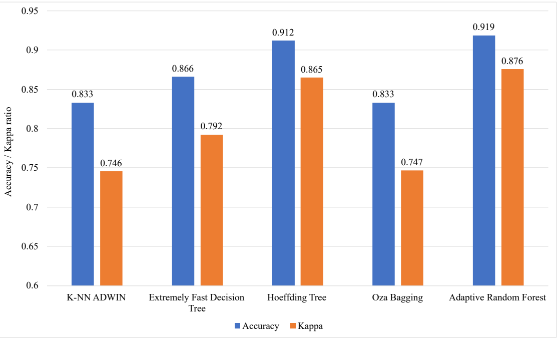


Fig. 14. Classification Performance Comparison among Five Adaptive Classifiers using Covid19-IQ02.

VI. Discussion

The COVID-19 pandemic has caused huge pressure on health systems in most countries due to a large number of infections. Thus, the focus of health institutions was to treat severe cases that needed breathing aids. Many patients were getting medical care at home in moderate and mild cases. Remote monitoring of the patient's recovery can reduce the necessary medical effort and speed up the return of the patient's normal life. This work proposes a classification system that determines whether or not a patient has recovered based on a set of patient biomedical and clinical data using machine learning techniques.

The process of collecting data on COVID-19 patients was one of the most difficult stages of this work due to the risks of transmission of the virus and continuous exposure to the virus. In addition, the quarantine laws for COVID-19 patients make them only accessible by certified health workers. Moreover, there was a challenge in finding COVID-19 patients health workers that are willing to volunteer for the study. This is due to the fact that volunteers require insurance and convincing that the proposed system guarantees security, ease of use, confidentiality, and effectiveness of using such health systems. In addition, due to the difficulty of gaining permission from governmental health intuitions to conduct this study, the focus was on private (non-governmental) medical clinics that were receiving patients, and they asked to use the bracelet to record their vital signs without using their names and any personal information using the proposed system.

The analysis of the collected data indicates that the vital signs data from the proposed wearable device sensors was more useful for the classification of the patients without chronic diseases. The patient's age significantly correlated with the recovery status, as the results showed. The younger patient tends to recover more than the adult and the elderly. However, a child or young person with a chronic disease is less likely to recover. Also, the impact of diabetes and its importance in building the classifiers, thereby predicting the recovery, was more than the impact of high blood pressure. However, the ranking of features during the process of building the classifier showed that the sensors data had more Gain ratio in the second dataset, which contains multiple readings (five) for each patient. In J48 classifier, which was the base learner for the two implementations, the root node was body temperature reading, and that can clarify the effectiveness of the proposed wearable device. Finally, as far as we know, this work is the first research analysis of COVID-19 patients' data using machine learning in Iraq. We look forward to extending the patients' sample and applying more improvements in data classification.

VII. Conclusion

Monitoring and home medical care contribute to focusing the medical effort on severe cases, especially during the spread of pandemics. The proposed system aims to design and implement a classifier capable of predicting a patient's recovery from COVID-19 and providing the medical staff with an immediate alert if the patient's condition declines. The implementation phase included collecting data from 408 patients, and the analysis of these data showed a significant correlation between the factors of chronic diseases, age, and patient recovery. The ensemble classification produced two classifiers. The first one was based on AdaBoost; it had the best accuracy of 0.874 compared with four classifiers with the first dataset and 0.998 with the second one. The second classifier was based on Adaptive Random Forest, which had the best accuracy of 0.919 compared with four adaptive classifiers. The ranking of features in AdaBoost classifiers showed more importance for the vital signs collected by the proposed system than the symptoms. Future works might include testing the the proposed classifier system on stream of data and provide instant predictions.

Acknowledgment

The authors would like to acknowledge the efforts of the medical staff in the private clinic of Dr. Abdullah Al-Rawiy in Alghazalyah in Baghdad for their efforts in collecting the data set.

References

[1] D. A. Schwartz and A. L. Graham, "Potential maternal and infant outcomes from coronavirus 2019-ncov (sars-cov-2) infecting pregnant women: lessons from sars, mers, and other human coronavirus infections," *Viruses*, vol. 12, no. 2, p. 194, 2020.

[2] Y. Zhu, Z. Wang, Y. Zhou, K. Onoda, H. Maruyama, C. Hu, and Z. Liu, "Summary of respiratory rehabilitation and physical therapy guidelines for patients with covid-19 based on recommendations of world confederation for physical therapy and national association of physical therapy," *Journal of physical therapy science*, vol. 32, no. 8, pp. 545–549, 2020.

[3] H. K. Naji, N. Goga, A. J. M. Karkar, I. Marin, and H. A. Ali, "Internet of things and health care in pandemic covid -19: System requirments evaluation," in *2021 7th International Engineering Conference "Research & Innovation amid Global Pandemic" (IEC)*, 2021, pp. 37–42.

[4] V. L. Yu and S. C. Edberg, "Global infectious diseases and epidemiology network (gideon): a world wide web-based program for diagnosis and informatics in infectious diseases," *Clinical infectious diseases*, vol. 40, no. 1, pp. 123–126, 2005.

[5] T. A. Morris, P. C. Gay, N. R. MacIntyre, D. R. Hess, S. K. Hanneman, J. P. Lamberti, D. E. Doherty, L. Chang, and M. A. Seckel, "Respiratory compromise as a new paradigm for the care of vulnerable hospitalized patients," *Respiratory care*, vol. 62, no. 4, pp. 497–512, 2017.

[6] L. Bouadma, F.-X. Lescure, J.-C. Lucet, Y. Yazdanpanah, and J.-F. Timsit, "Severe sars-cov-2 infections: practical considerations and management strategy for intensivists," *Intensive care medicine*, vol. 46, no. 4, pp. 579–582, 2020.

[7] R. T. Gandhi, J. B. Lynch, and C. Del Rio, "Mild or moderate covid-19," *New England Journal of Medicine*, vol. 383, no. 18, pp. 1757–1766, 2020.

[8] N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, and H. Nishiura, "Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data," *Journal of clinical medicine*, vol. 9, no. 2, p. 538, 2020.

[9] H. K. Naji, N. Goga, A. J. M. Karkar, H. A. Ali, and M. Falahi, "A prototype to monitor vital signs, locate, and track covid-19 patients in quarantine zones," in *5th International Conference on Computer Applications and Information Security (ICCAIS'2022), Accepted*, 2022.

[10] M. Mohammed, S. Desyansah, S. Al-Zubaidi, and E. Yusuf, "An internet of things-based smart homes and healthcare monitoring and management system," in *Journal of Physics: Conference Series*, vol. 1450, no. 1. IOP Publishing, 2020, p. 012079.

[11] B. De Ville and P. Neville, *Decision trees for analytics: using SAS Enterprise miner*. SAS Institute Cary, NC, 2013.

[12] N. Al Bassam, S. A. Hussain, A. Al Qaraghuli, J. Khan, E. Sumesh, and V. Lavanya, "Iot based wearable device to monitor the signs of quarantined remote patients of covid-19," *Informatics in medicine unlocked*, vol. 24, p. 100588, 2021.

[13] M. D. Santos, C. Roman, M. A. Pimentel, S. Vollam, C. Areia, L. Young, P. Watkinson, and L. Tarassenko, “A real-time wearable system for monitoring vital signs of covid-19 patients in a hospital setting,” *Frontiers in Digital Health*, vol. 3, 2021.

[14] M. M. Rahma and A. D. Salman, “A wearable medical monitoring and alert system of covid-19 patients,” *Iraqi Journal for Computers and Informatics*, vol. 47, no. 1, pp. 12–17, 2021.

[15] G. Cosoli, S. Spinsante, and L. Scalise, “Wrist-worn and chest-strap wearable devices: Systematic review on accuracy and metrological characteristics,” *Measurement*, vol. 159, p. 107789, 2020.

[16] A. F. Al-zubidi, N. F. AL-Bakri, R. K. Hasoun, S. H. Hashim, and H. T. Alrikabi, “Mobile application to detect covid-19 pandemic by using classification techniques: Proposed system.” *International Journal of Interactive Mobile Technologies*, vol. 15, no. 16, 2021.

[17] S. Sakib, T. Tazrin, M. M. Fouda, Z. M. Fadlullah, and M. Guizani, “Dl-crc: deep learning-based chest radiograph classification for covid-19 detection: a novel approach,” *Ieee Access*, vol. 8, pp. 171 575–171 589, 2020.

[18] M. Pahar, M. Klopper, R. Warren, and T. Niesler, “Covid-19 cough classification using machine learning and global smartphone recordings,” *Computers in Biology and Medicine*, vol. 135, p. 104572, 2021.

[19] M. Polsinelli, L. Cinque, and G. Placidi, “A light cnn for detecting covid-19 from ct scans of the chest,” *Pattern recognition letters*, vol. 140, pp. 95–100, 2020.

[20] H. K. Naji, N. Goga, A. J. M. Karkar, H. A. Ali, M. Falahi, and A. Al-Rawiy, “Design and development of a bracelet to monitor vital signs of covid-19 patients,” in *5th International Conference on Engineering Technology and its Applications 2022- (5thIICETA2022)), Accepted*, 2022.

[21] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[22] H. M. Gomes, A. Bifet, J. Read, J. P. Barddal, F. Enembreck, B. Pfharinger, G. Holmes, and T. Abdessalem, “Adaptive random forests for evolving data stream classification,” *Machine Learning*, vol. 106, no. 9, pp. 1469–1495, 2017.

[23] D. Chen, F. Song, L. Tang, H. Zhang, J. Shao, R. Qiu, X. Wang, and Z. Ye, “Quarantine experience of close contacts of covid-19 patients in china: a qualitative descriptive study,” *General hospital psychiatry*, vol. 66, pp. 81–88, 2020.

[24] “Rapidminer,” accessed: 2022-05-16. [Online]. Available: https://rapidminer.com/

[25] “Waikato environment for knowledge analysis,” accessed: 2022-05-12. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/

Mobile Application Design: Sale of Clothes Through Electronic Commerce

Raul Jauregui-Velarde[1], Franco Gonzalo Conde Arias[2], Jose Luis Herrera Salazar[3],
Michael Cabanillas-Carbonell[4], Laberiano Andrade-Arenas[5]
Facultad de Ingeniería y Negocios
Universidad Privada Norbert Wiener
Lima, Perú[1,2,3,5]
Facultad de Ingeniería
Universidad Privada del Norte
Lima, Perú[4]

Abstract—**During the COVID-19 pandemic, small clothing sales companies lost economic income and customers due to a lack of digital transformation, causing the dismissal of many employees. Due to this problem, our objective is to design an e-commerce mobile application for the sale of clothes, so that Small and medium-sized enterprises dedicated to this area generate income and retain their customers. For this, the Rational Unified Process (RUP) methodology was applied, because this methodology provides a structured way for companies or developers to visualize the development of the software, and for the validation by expert judgment, the survey and the questionnaire were used as instruments. Obtaining as a result a positive rating for the design of the mobile application and its acceptance to accommodate what is reflected. In conclusion, the e-commerce mobile application was successfully designed, backed by expert judgment, so that Small and medium-sized enterprises can offer their products and generate income, as well as build customer loyalty.**

Keywords—Mobile application; COVID-19; e-commerce; RUP; sale of clothes

I. Introduction

At present, worldwide, mobile applications have become elements of vital importance for electronic commerce. Since, during the pandemic, many companies have gone downhill, especially small companies due to lack of financial resources for digital transformation. Large companies, however, showed a slightly different result compared to small companies. This is due to the use of smartphones, which has generated a lot of innovation and with it, the creation of mobile applications, including the e-commerce application [1].

During the COVID-19 pandemic, mobile applications for online commerce or also called electronic commerce, gained a large number of new users, and in the same way, people's lifestyles have changed [2]. Therefore, today consumers choose to use these applications to make purchases from the comfort of their home. Because mobile applications have emerged as a form of innovation for the e-commerce business, since it provides convenience and ease to consumers by saving time and effort. In this sense, people prefer the use of electronic commerce applications, in particular to make purchases or orders, in this way, protect themselves from the COVID-19 coronavirus [3].

In addition, mobile e-commerce applications can help small businesses that sell clothes, and other sectors that have the home delivery modality, prosper. In the same way, it will allow consumers (customers) to make their purchases from the comfort and safety of their homes [4]. According to the study carried out in India, the habits adopted during the COVID-19 coronavirus pandemic are driving the growth of mobile applications related to electronic commerce. Because these applications help save time, resources and efforts for consumers. Therefore, it is one of the key reasons why small businesses can prosper [5]. Likewise, the use of the mobile application generates great benefits for the company, such as: quick and easy access to information, optimization of the average time in making and delivering purchases or orders and, consequently, improvement in customer loyalty.

For this reason, the present research work offers a viable solution to small companies dedicated to the sale of clothes, using an electronic commerce mobile application. In order to help small businesses prosper in the current market, to which they are directed.

Therefore, the proposed research will help automate and provide a solution to the problem of face-to-face clothing sales that currently exists, allowing customers to order or purchase with confidence and security from the mobile application.In this way, allow small businesses to provide services at home. Since the mobile application is key to continue generating income and sustain itself in the current market. Likewise, with this form of technological innovation, it not only benefits customers, but also small businesses and people who work as delivery people, since the mobile application is a system that helps the efficiency of the business and involves several people. that require a job. In the same way, the RUP methodology will be applied since this methodology provides a structured way to visualize software development. It also provides a detailed plan for each of the development phases.

The objective of the research is to design a mobile application for the sale of clothes through electronic commerce, applying the RUP methodology, which makes it easier for small businesses to generate income and retain their customers. The beneficiaries of the research are the people who sell clothing to generate income through the proposal of the mobile application, all through electronic commerce.

This research is made up as follows: Section II explains the review of the literature: analysis of different investigations related to the research work; Section III defines the RUP methodology and its different phases that were used in the research work; Section IV the development of the methodology; Section V presents the results; Section VI the discussions; Section VII the conclusions and finally Section VIII future work.

II. Literature Review

In the research work, the subject of e-commerce mobile applications was addressed. For this reason, in this section it will focus on analyzing the different investigations related to the research work, where it provides us with its objectives, methods, results and conclusions.

The author Hawa [6], states that the internet, applications and e-commerce allow consumers to be reached, since these mobiles allow the availability of clothing; therefore, sales increase. He developed the study with the aim of exploring the acceptance of personalizing clothing online, with the purpose of reaching the market through segmentation. Applying the qualitative method; the interview and questionnaire for data collection. The sample consists of 13 participants from diverse cultural backgrounds. Their findings contribute to a large-scale quantitative study. In addition, it is of interest to great executives of management and marketing of garments.

The author M.Subchan [7], manifest that electronic commerce allows consumers to carry out purchase and sale transactions in a simple and fast way. Likewise, it presents that the Muslim clothing store ayu fashion shops, offers a variety of Muslim clothing such as: hijab, tunic and couple; therefore, the purpose of the study is to create a sales system using electronic commerce technology, applying the waterfall methodology. Concluding that consumers of these online clothing stores can order from anywhere. Likewise, it provides the solution to the problems faced by Muslim fashion stores that are currently evolving.

The authors Purwaningtyas and Rahadi [8], in their study, aim to determine the factors that affect the sale of clothing through electronic commerce; To do this, they reviewed and came to a synthesis of 36 previous articles related to their research. Obtaining a result, that the factors that affect the sale of clothing through electronic commerce are the price, the design and style of the product, the promotion, the quality of the product, the availability of product information, the variety of products offered, ease of use and quality of service during the purchase. The study's findings are of great use to businesses selling garments on e-commerce platforms in Jakarta, Indonesia.

The authors BĂLĂȘESCU et al. [9], state that the development of electronic commerce platforms is of vital importance for the clothing sales market. The study analyzes clothing sales via e-commerce in Romania; For this, it is intended to know the opinions of clothing consumers, who use electronic commerce applications to make their purchases. In this way, determine the reasons why they choose to buy clothes online. Concluding that the results obtained in the study will be very useful for online clothing sales companies through applications, which will help improve their services. Likewise, for companies that want to enter the e-commerce business through their applications.

The authors Ramirez et al. [10], in the investigation present the development of an electronic commerce platform, to give the greatest flow of sales of sports and casual clothing. Similarly, they state that during the COVID-19 pandemic, many clothing sales businesses were affected; For this reason, the electronic commerce platform was developed, applying the Scrum methodology to solve the problem. According to the statistics they obtained, the acceptance by the clients was evidenced, since, from the moment it was put into service, the business obtained positive results. Concluding that e-commerce applications increase sales by making it easier for consumers to buy online.

The authors Soegoto et al. [11], state that there is a rapid development of technologies; one of them is the development of mobile applications as a means for the sale and purchase of a product. The purpose of the research is to design a mobile application for the sale and purchase of various types of Japanese anime t-shirts. Also, as a means of promotional marketing. For this, the qualitative descriptive analysis method was applied, in order to help Japanese anime fans to find different types of t-shirts. Likewise, it concludes that the mobile application facilitates the purchase and acts as an intermediary between consumers and the favorite anime product.

The authors Gomero-Fanny et al. [12] in their research work made a prototype of electronic commerce for the sale of clothes. To do this, they applied the scrum methodology. The objective was to design a web system for clothing sales under the agile SCRUM methodology. Which allowed them to design web system prototypes meeting the needs of the organization. The results were divided into 4 Sprint deliverables to analyze the user stories, where a maximum score of 21 and a minimum of 16 were obtained. Concluding that this system will focus on meeting the requirements of customers, with adaptations to the organization according to their needs, which will benefit the organization and its customers.

So also the authors Tupia-Astoray and Andrade-Arenas [13], state that currently Small and medium-sized enterprises (SMEs) have stalls and physical stores as the only means of sale. For this reason, in their study, e-commerce web prototypes for sale were made, applying the SCRUM methodology covering the established procedures, being a peculiar proposal and with a beneficial approach. Obtaining as a result, design of innovative prototypes complying with the procedures established under the SCRUM methodology. Therefore, the proposal made can be implemented by different SMEs that wish to improve their online sales, with a good management organization.

The authors Lazo-Amado et al. [14] mention in their research work that the COVID-19 pandemic generated a great loss of sales in the Peruvian market. For this reason, the objective of the study is to develop a model to optimize sales with the use of digital marketing, applying the DesignScrum methodology, which is a hybrid of Scrum and Design Thinking. To carry out the test, a survey was conducted with customers, who gave their opinion regarding the prototype. Then, the digital marketing proposal was raised. Concluding that the marketing model according to the needs of the company will benefit its sales through electronic business.

In summary, different research works have been studied and it was found that most of the authors focus on the development of web applications with attributes such as quality, design, reliability, price and various options associated with web applications. applications, and others focuses on the factors that influence customers to use the mobile application. However, they do not focus on the development of the mobile application, on the security that it can provide to the user. Likewise, they do not attribute the forms of payment to facilitate the consumer and can use them with confidence and security. Consequently, they are exposed to not achieving customer satisfaction and loyalty.

III. Methodology

This section focus on applying the RUP methodology, which is made up of phases.

A. Metodología RUP

The RUP methodology is a software development process that divides the process into four distinct phases, such as: start, build, build, and transition [15]. The four phases are as follows (see Fig. 1) :

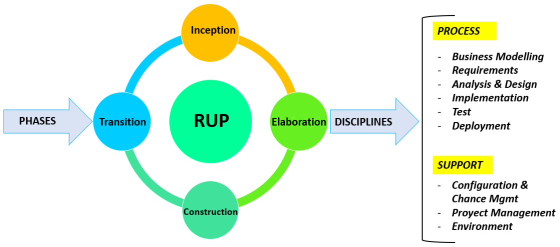


Fig. 1. Phases of the RUP Methodology.

1) Start: in this initiation phase, the project is defined. Likewise, it is determined if the project is feasible. In addition, the vision and objectives are defined, as well as the scope of the project. Similarly, it focuses on business modeling and identifying its requirements.

2) Elaboration: in this phase, the use cases are selected, which will allow defining the architecture of the system to be developed. Likewise, we proceed with the specification of each of the use cases and the analysis of the domain is carried out. In the same way, the design of the preliminary solution is carried out.

3) Construction: In this phase, the construction or coding of the software is carried out, which is carried out following a series of iterations. Afterwards, it begins with its implementation and proceeds with its respective tests, in order to find flaws or defects in the created software.

4) Transition: In this phase, the construction or coding of the software is carried out, which is carried out following a series of iterations. Afterwards, it begins with its implementation and proceeds with its respective tests, in order to find flaws or defects in the created software.

B. Design Tool

In this section, the design tools for the prototype of the mobile application were detailed.

1) Figma: In this project, the use of the Figma tool was used, an intuitive cloud-based application that allows designers to work without previously downloading the software [16]. It also allows the contribution of a work team in real time.

2) StarUML: This tool is developed under the Unified Modeling Language (UML) and Modeler Driven Architecture (MDA) standards, which allows modeling diagrams required for development and implementation in software projects. In the same way, it allows to obtain a better view of their operation [17].

C. Development Tools

In this section, the mobile application development tools were detailed.

1) Kotlin: Modern programming language, it is statically typed, capable of running on the Java Virtual Machine (JVM) platform. Likewise, it presents great advantages to the developer, such as: the reduction of the level of complexity of the code that is usually written during the development of the mobile application. In addition, it contains a large number of material design components that can be used to support UI/UX related interfaces [18]. Similarly, it is compatible with the java language, thanks to NativeScript.

2) SQLite: The chosen database engine is SQLite, developed under the sql language. This database manager is fast, highly reliable, self-contained, and small with big features [19]. Therefore, it allows you to store in a simple, fast and efficient way. Also, the implementation of this is very simple and light. Similarly, it is robust and totally free.

3) Android Studio IDE: Android Studio is free and open source software. Likewise, it is the software most used by developers for the development of Android mobile applications. In addition, it provides developers with the fastest tools to create applications for all types of Android devices. Also, the editing of the source code is world class. Since, it features debugging, performance tools, a very flexible build system. Which allow the developer to create unique and high-quality applications [20]. With this, scalable projects can be carried out quickly and efficiently.

IV. Development of the Methodology

In this section, the business model and the system model were developed to develop the mobile application.

A. Business Modeling

1) Current Business Process: Fig. 2 shows the current business model, which is working without the proposed mobile application.

2) Business use case Modeling:

- Business actor: Fig. 3 shows the actors of the business.
- Business use case diagram: Fig. 4 the business use cases.

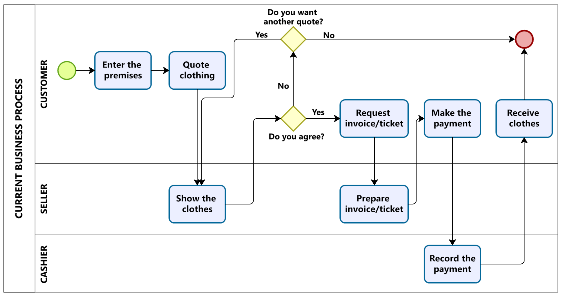


Fig. 2. Current Company Process.

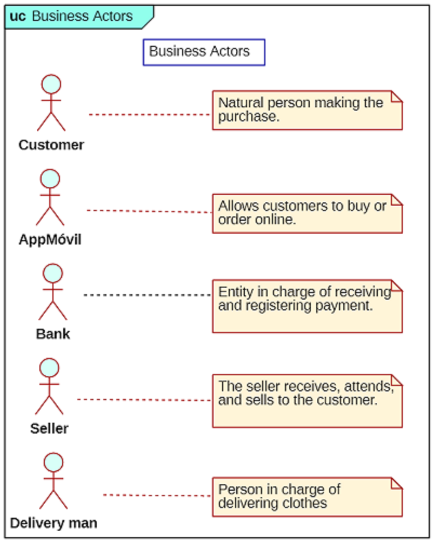

uc Business Actors
Business Actors
Customer
Natural person making the purchase.
AppMóvil
Allows customers to buy or order online.
Bank
Entity in charge of receiving and registering payment.
Seller
The seller receives, attends, and sells to the customer.
Delivery man
Person in charge of delivering clothes

Fig. 3. Business Actors.

3) Business Analysis Model:

- Business Sequence Diagram: Fig. 5 shows the sequence of business clothing sales steps.

B. Requirements Capture

- Functional Requirements: Table I shows the functional requirements of the mobile application for the sale of clothes.

TABLE I. FUNCTIONAL REQUIREMENTS

N°	Functional Requirements
1	User register
2	User authentication
3	List of products
4	Validate user
5	Search product
6	add to cart
7	generate payment
8	Register order
9	Search order
10	Cancel an order
11	Search category
12	Customer maintenance
13	Remove product from cart
14	Generate electronic bill
15	Print invoice

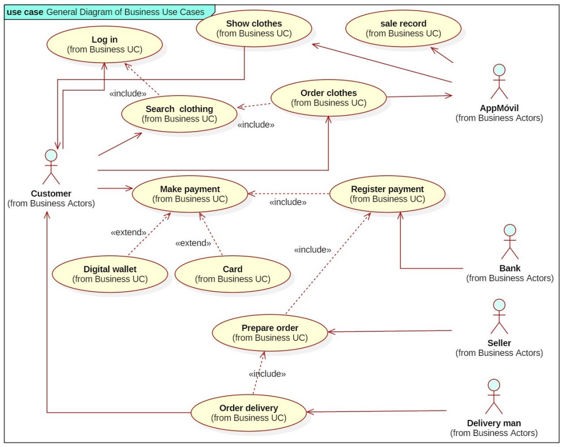


Fig. 4. Business Use Cases.

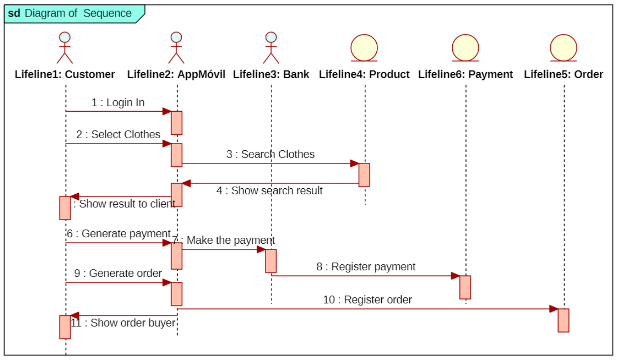


Fig. 5. Business Sequence Diagram.

- Non-Functional Requirements: Table II shows the non-functional requirements of the mobile application for the sale of clothes.

TABLE II. NON-FUNCTIONAL REQUIREMENT

N°	Non-Functional Requirement
1	The response time for order search is no more than 1 second.
2	Friendly, dynamic, descriptive and informative interface
3	Payment options
6	The system will have to be updated by users
5	The graphical interface of the system must be easy to read for the user
6	Allow remove product added from cart
7	Track purchase order with location in real time.
8	Request verification code when creating the account.
9	Request the password when deleting the account.
10	When creating the account you must accept the privacy terms
11	You must have the option to register with google.
12	You should have the option to log in without registering.

C. Modeling of System Use Cases

- System actor: Fig. 6 shows the actor that will interact with the mobile application.

- System use cases: Fig. 7 shows the use cases of the system.

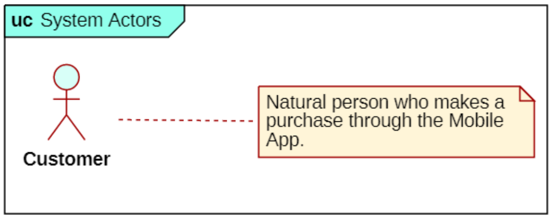


Fig. 6. System Actor.

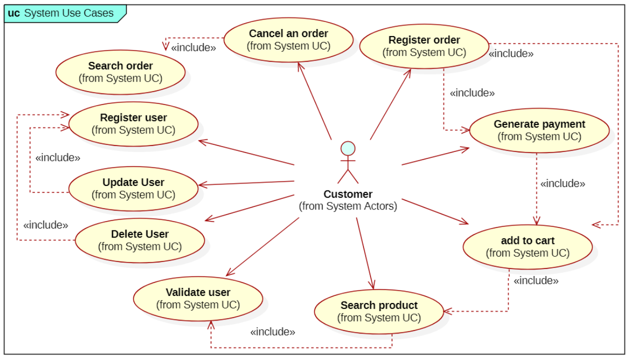

uc System Use Cases
Cancel an order (from System UC)
Register order (from System UC)
Search order (from System UC)
Register user (from System UC)
Update User (from System UC)
Delete User (from System UC)
Validate user (from System UC)
Search product (from System UC)
add to cart (from System UC)
Generate payment (from System UC)
Customer (from System Actors)
«include»
«include»
«include»
«include»
«include»
«include»
«include»
«include»

Fig. 7. System Use Cases.

D. System Analysis Modeling

- System activity diagram: Fig. 8 shows the sequence of activities that the client must carry out to register or buy the clothes through the mobile application.

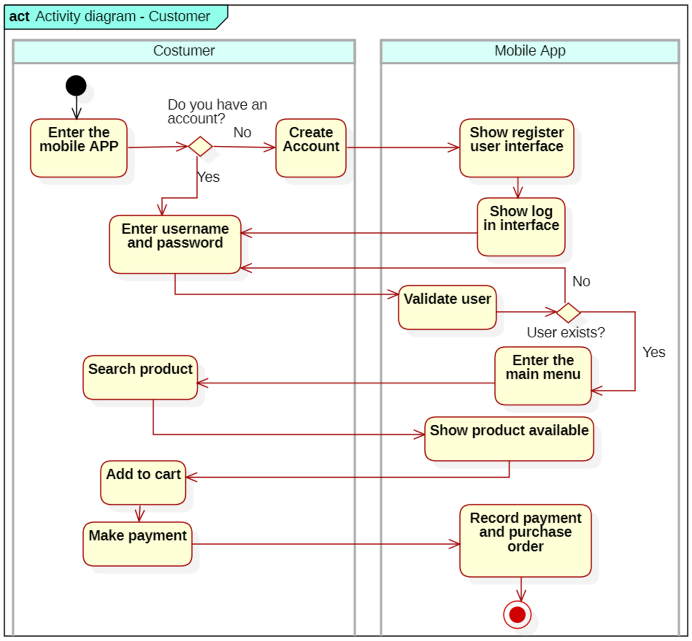


Fig. 8. System Activity Diagram.

- System Sequence Diagram - Register Order: Fig. 9 shows the order registration sequence.
- System sequence diagram - user

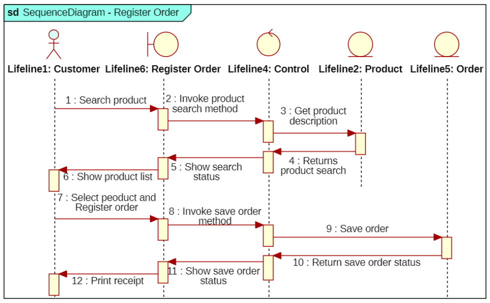

sd SequenceDiagram - Register Order
Lifeline1: Customer
Lifeline6: Register Order
Lifeline4: Control
Lifeline2: Product
Lifeline5: Order
1 : Search product
2 : Invoke product search method
3 : Get product description
4 : Returns product search
5 : Show search status
6 : Show product list
7 : Select peoduct and Register order
8 : Invoke save order method
9 : Save order
10 : Return save order status
11 : Show save order status
12 : Print receipt

Fig. 9. System Sequence Diagram - Register Order.

- ○ Register User: Fig. 10 shows the user registration sequence in the mobile application.

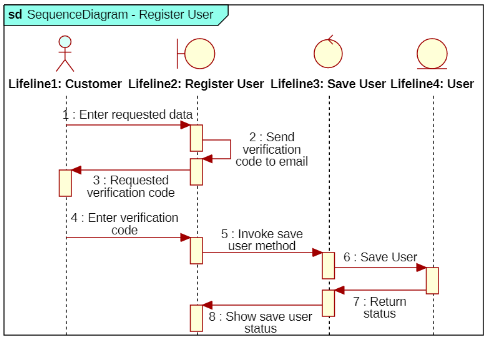

sd SequenceDiagram - Register User
Lifeline1: Customer
Lifeline2: Register User
Lifeline3: Save User
Lifeline4: User
1 : Enter requested data
2 : Send verification code to email
3 : Requested verification code
4 : Enter verification code
5 : Invoke save user method
6 : Save User
7 : Return status
8 : Show save user status

Fig. 10. System Sequence Diagram - Register User.

- ○ Update User: Fig. 11 shows the user data update sequence in the mobile application.

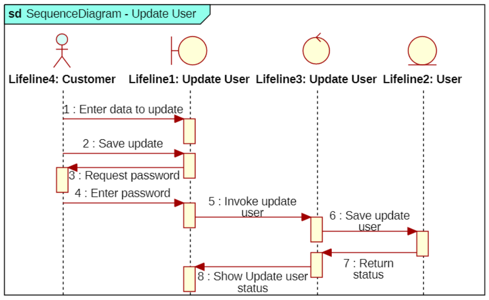


Fig. 11. System Sequence Diagram - Update User.

- ○ Delete user: Fig. 12 shows the sequence for deleting the user account in the mobile application.

- Entity-Relationship Diagram: Fig. 13 shows relationship of the database entities.

E. Construction of the mobile application

In this section you will focus on mobile application design.

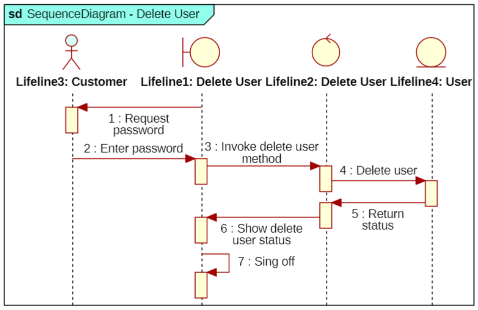
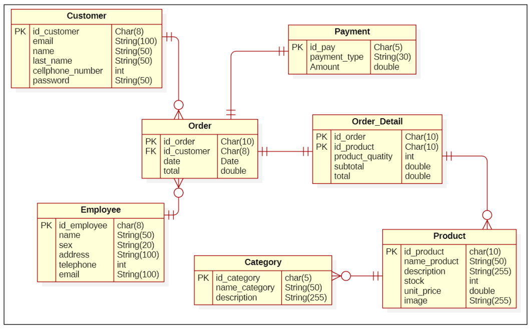


Fig. 12. System Sequence Diagram - Delete User.



	Customer	
PK	id_customer	Char(8)
	email	String(100)
	name	String(50)
	last_name	String(50)
	cellphone_number	int
	password	String(50)

	Payment	
PK	id_pay	Char(5)
	payment_type	String(30)
	Amount	double

	Order	
PK	id_order	Char(10)
FK	id_customer	Char(8)
	date	Date
	total	double

	Order_Detail	
PK	id_order	Char(10)
PK	id_product	Char(10)
	product_quatity	int
	subtotal	double
	total	double

	Employee	
PK	id_employee	char(8)
	name	String(50)
	sex	String(20)
	address	String(100)
	telephone	int
	email	String(100)

	Category	
PK	id_category	char(5)
	name_category	String(50)
	description	String(255)

	Product	
PK	id_product	char(10)
	name_product	String(50)
	description	String(255)
	stock	int
	unit_price	double
	image	String(255)

Fig. 13. Entity Relationship Diagram.

V. Results

A. About the Prototype

Fig. 14 presents the login and user registration modules. The registration module allows you to create an account and by creating it, the user gets access to the virtual store. Similarly, the login module allows the user to validate their data for security reasons and make purchases with confidence.

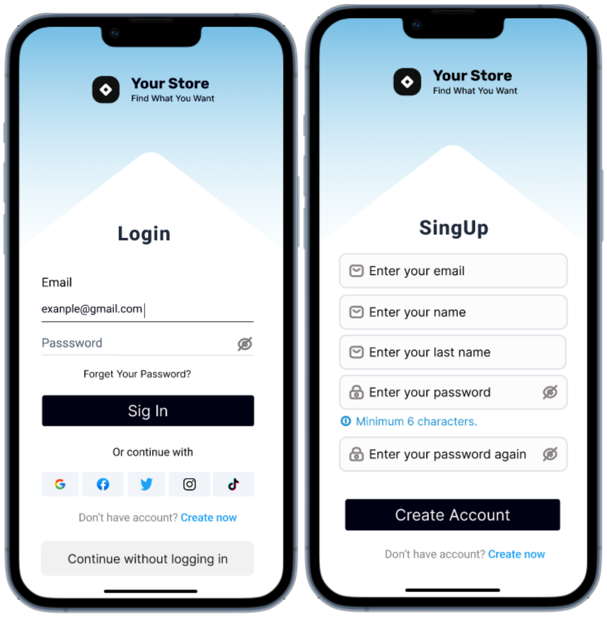


Fig. 14. Sign in and Sign up Prototype.

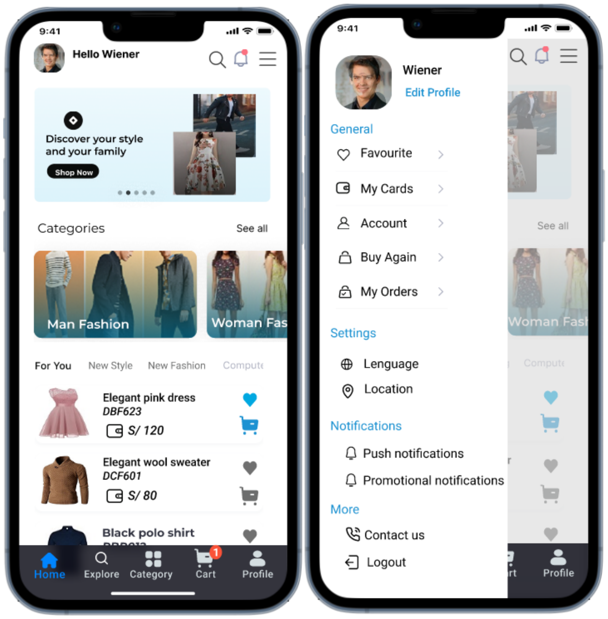


Fig. 15. Prototype Main Menu and Navigation Menu.

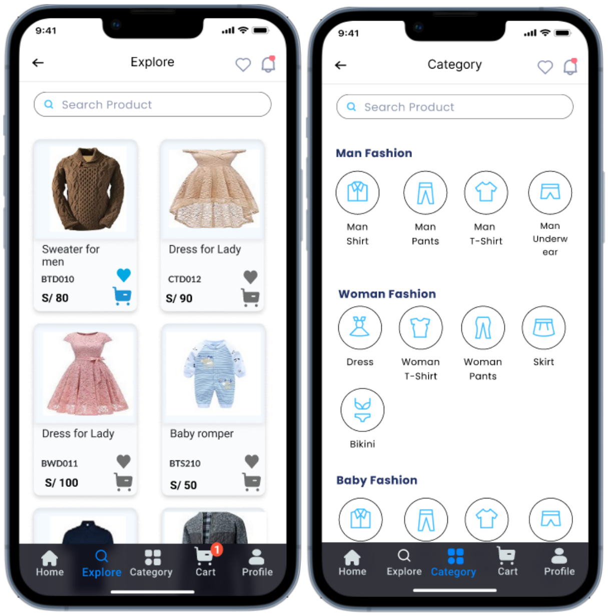


Fig. 16. Prototype Explore and Category.

In Fig. 15, the main menu and the navigation modules are presented. The main menu module shows the recently published products (clothes) with their respective prices and their shopping cart button to facilitate the user's choice. Likewise, the navigation menu module makes it easier for the user to navigate in the application by presenting different options in a simple way.

In Fig. 16, the exploration and category modules were presented. The explore module makes it easy for the user to search for the clothes they want and in the same way the module filters the product with its respective prices according to what the user is looking for. Likewise, the category module

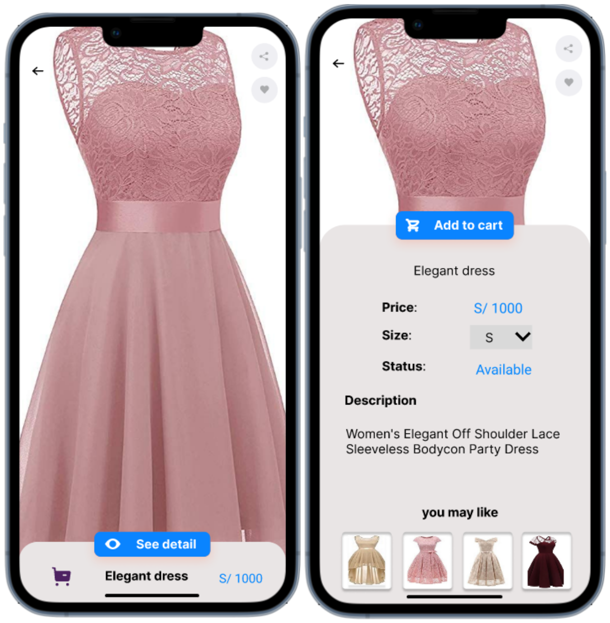


Fig. 17. Prototype to See and See in Detail.

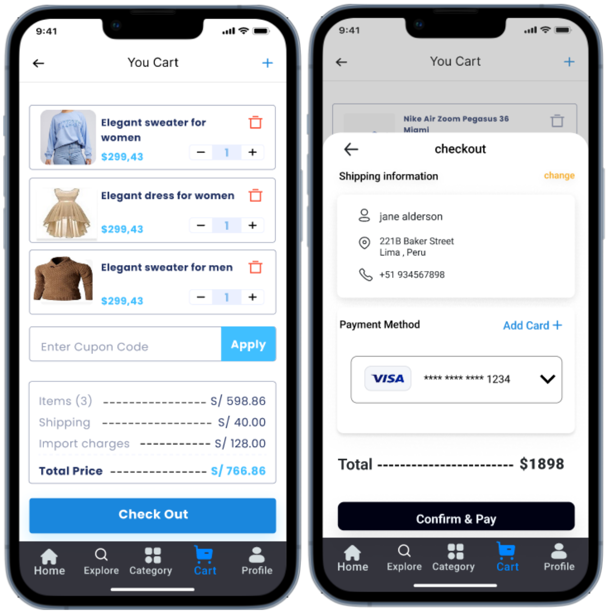


Fig. 18. Shopping Cart and Checkout Prototype.

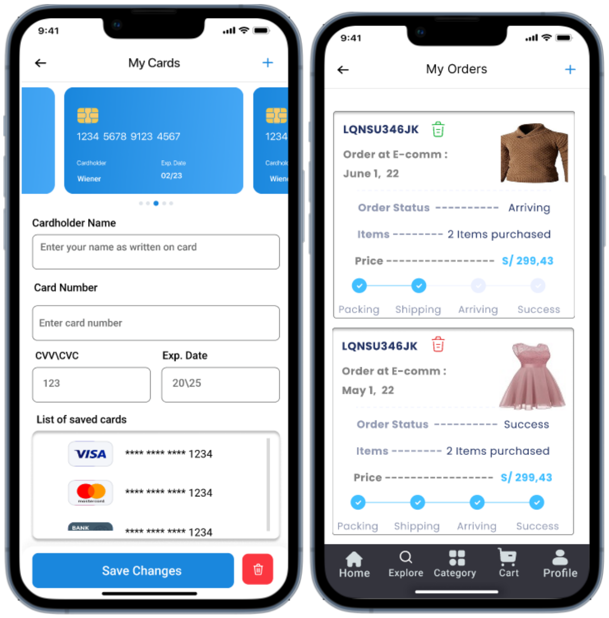


Fig. 19. Prototype of Cards and Purchase Order

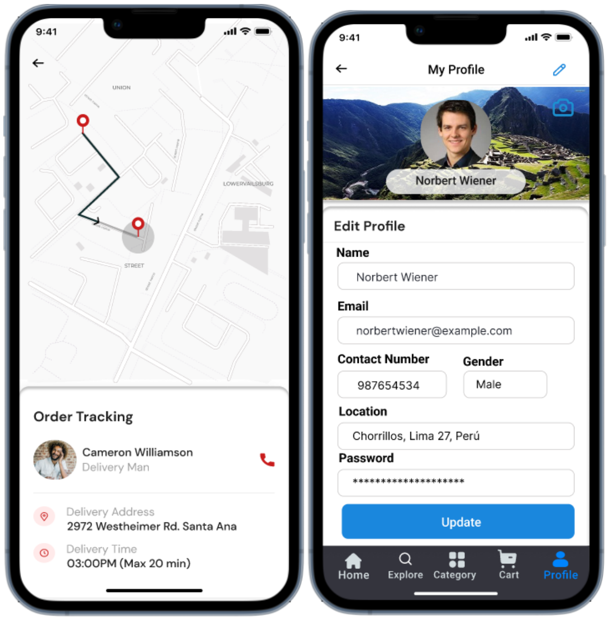


Fig. 20. Order Tracking and User Profile Prototype.

allows the user to search by category, classifying by gender and age.

In Fig. 17, the product view and detail view modules were presented. The product view module allows the user to view the clothing from different angles. Similarly, the detailed view module allows the user to review the information related to the product and make a decision when making the purchase.

In Fig. 18, the modules my cart and verify purchase were presented. The my cart module shows the user the summary or detail of their purchase such as: clothes to buy, quantity to buy, shipping cost, subtotal and total purchase. Similarly, the purchase verification module shows the information of the purchasing user, such as: name, delivery address and contact number. In addition, the form of payment and the total of the purchase to be confirmed.

In Fig. 19, the modules my cards and my orders were presented. The my card module allows the user to add their cards with which they can make payments for their purchases safely. On the other hand, the my orders module shows the status of the purchase and allows the user to track their purchase.

Fig. 20 shows the purchase tracking modules and the user profile. The purchase tracking module allows the user to follow their purchase through a map, which shows the location of the personnel in charge of delivering the purchase; as well as staff information, contact number and delivery time. While the profile module shows the user information. It also allows you to update your data.

B. About the Survey

For the survey, an evaluation questionnaire of 5 criteria or dimensions was elaborated, each one with 4 questions as shown in Table III. In which the evaluation of 10 experts was requested on the following criteria that the mobile application must meet: the design, usability, functionality, security and availability of the mobile application.

Regarding the items or questions for the evaluation, the three-point measurement scale was determined, in which Low, Medium and High must be evaluated according to the established criteria.

TABLE III. Evaluation Questions

Criterion	**Design**
P1	¿Does the application present a simple and neat user interface?
P2	¿Does the application present classified products?
P3	¿Does the application have simplified and clear navigation?
P4	¿Does the application present valuable content?
Criterion	**Usability**
P5	¿Is the application easy to understand and intuitive?
P6	¿Is the app easy to learn to use its features?
P7	¿Is ordering done quickly and easily in the app?
P8	¿Is the app easy and simple to navigate?
Criterion	**Functionality**
P9	Does the application have a simple and fast payment process?
P10	¿Does the application allow you to add or remove products from the shopping cart?
P11	. ¿Does the application filter products according to the searched product?
P12	¿Does the application allow the customer to have control of their order?
Criterion	**Security**
P13	¿Does the application request user authentication to make purchases?
P14	¿Does the application guarantee the integrity of customer data?
P15	¿Does the application maintain the availability of customer data?
P16	¿Does the application present a secure payment process and methods?
Criterion	**Availability**
P17	¿In the application the shopping cart is always visible?
P18	¿Does the application have payment methods commonly used in your online purchases?
P19	¿In the application is the description of the shipping options?
P20	¿Does the app show available products?

C. About Expert Evaluation

After evaluating the experts and analyzing their answers, the mean and standard deviation of each question were obtained. Also, according to the mean of the question, it was scored on a scale, which includes the following: 0 – 1 = Bass, 1.1 – 2 = Medium, 2.1 – 3 = High. Obtaining a result as shown in Table IV.

According to the results, in the Design criterion, in question 1, does the application present a simple and orderly user interface? An average of 3.00 or also called average was obtained, with a standard deviation of 0.000 to which the experts gave the average as High. Likewise, in the Usability criterion, in question 5, is the application easy to understand and intuitive? A mean of 2.70 with a standard deviation of 0.483 was obtained, to which the experts gave the mean as High.

TABLE IV. Result of the Evaluation of the Experts

Criterion	**Question**	**Half**	**SD**	**Scale**
Design	P1	3,00	0,000	High
	P2	2,90	0,316	High
	P3	2,90	0,316	High
	P4	2,90	0,316	High
Usability	P5	2,70	0,483	High
	P6	3,00	0,000	High
	P7	2,80	0,422	High
	P8	3,00	0,000	High
Functionality	P9	2,80	0,422	High
	P10	3,00	0,000	High
	P11	2,90	0,316	High
	P12	3,00	0,000	High
Availability	P13	3,00	0,000	High
	P14	3,00	0,000	High
	P15	3,00	0,000	High
	P16	3,00	0,000	High
Security	P17	3,00	0,000	High
	P18	2,90	0,316	High
	P19	2,80	0,422	High
	P20	3,00	0,000	High

D. About the Methodology

In this section, the comparison between the RUP, Extreme Programming (XP) and Rapid Application Development (RAD) methodologies will be made.

For the evaluation of the methodology, numbers from 1 to 5 were used, where (1) indicates that the practices are unfavorable for the correct development of the project. For its part, (5) indicates that the analyzed criterion is the one with the best compliance with respect to the practices for the development of the project. Obtaining a result as shown in Table V.

According to the results obtained, the RUP methodology has a score of 26, the XP methodology obtains a score of 20, and the RAD methodology obtains a score of 13. Which means that the RUP methodology is the most appropriate to develop the project.

TABLE V. Quantitative Selection Criteria

Criterion	**RUP**	**XP**	**RAD**
Budget available	3	4	3
Project size	5	2	1
Limited delivery times	2	4	2
Need for documentation	5	2	1
Staff needed	5	2	1
Adaptability, response to changes	2	4	2
Customer Impossibility	4	2	3
TOTAL	26	20	13

VI. Discussion

In the findings found in the investigation, it was through a survey of the experts, who qualified as high on a three-point measurement scale; low, medium and high. However, the authors' research [14] is different from ours, since they applied a survey to customers, who gave their opinion. Regarding the prototype and the methodology, the authors [12] developed the clothing sales web system applying the agile methodology such as SCRUM, if we compare it with the research carried out, it does not coincide because ours is focused on the design of a mobile application under the traditional methodology that is the RUP. Regarding the segmentation, the authors' research

[11] applies the development of the mobile application only for the sale of Japanese anime t-shirts, which is only for a community that likes anime t-shirt. For its part, our research focuses on the development of the application for the sale of all types of clothing, which is segmented for men, women, boys, girls and adolescents from different cultures. Finally, in our work we came to the conclusion that the e-commerce mobile application for the sale of clothing can generate economic income and help retain customers, in the same way help grow small businesses that are growing. ; agreeing with the authors [7] who conclude that the mobile application allows the customer to make purchases from anywhere, in the same way it helps Muslim fashion stores that are evolving or growing.

VII. Conclusion

In conclusion, in the present research work, it was possible to design a mobile application for electronic commerce in a satisfactory manner by users, backed by expert judgment based on 5 criteria: design, usability, functionality, security and availability; which guarantee that the application is of quality, applying the RUP methodology. In the same way, the application makes it easier for small companies or SMEs dedicated to the sale of clothing, to offer their products through electronic commerce, thus improving their economic income and building customer loyalty. Likewise, the RUP methodology is a method that ensures that the development is of the best quality and foresees the changes that occur during the development according to the requirements. In the same way, it shows a global vision and optimizes its development.

A limitation of the investigation was the technological one, due to the retractions that the hardware imposes on us when designing the mobile application. Since, when designing the application, you should think about whether the application will be installed satisfactorily on a wide range of mobile devices, since some are high-end and others are not.

VIII. Future Work

In addition, it is suggested as future work that this work should be completed with emerging technologies such as, artificial intelligence integrating the application such as chatbots (virtual assistant) to interact with customers. Likewise, automated personalization, so that the mobile application has the capacity to give recommendations to the client according to their interactions and purchases they make.

References

[1] A. Jamin, I. N. Zukri, N. Yazid, N. Ahmad, and S. R. Sakarji, "The relationship between food delivery application (fdas) attributes and customers' satisfaction during covid-19," *International Journal of Accounting, Finance and Business*, vol. 6, 2021.

[2] W. Puriwat and S. Tripopsakul, "Understanding food delivery mobile application technology adoption: A utaut model integrating perceived fear of covid-19," *Emerging Science Journal*, vol. 5, 2021.

[3] T. Dirsehan and E. Cankat, "Role of mobile food-ordering applications in developing restaurants' brand satisfaction and loyalty in the pandemic period," *Journal of Retailing and Consumer Services*, vol. 62, p. 102608, 9 2021.

[4] D. Pal, S. Funilkul, W. Eamsinvattana, and S. Siyal, "Using online food delivery applications during the covid-19 lockdown period: What drives university students' satisfaction and loyalty?" *Journal of Foodservice Business Research*, 2021.

[5] R. Ramesh, S. V. Prabhu, B. Sasikumar, B. K. Devi, P. Prasath, and S. P. R. Kamala, "An empirical study of online food delivery services from applications perspective," *Materials Today: Proceedings*, 6 2021.

[6] H. Hawa, "Attitudes toward apparel mass customization: Canadian consumer segmented by lifestyle and demographics," vol. 113, 2018.

[7] D. S. M. Subchan, "Information system for sale of muslim clothes based on e-commerce technology," *Jurnal Mantik*, vol. 4, 2020.

[8] A. Purwaningtyas and R. A. Rahadi, "The affecting factors on online clothing purchase: A conceptual model," *Advanced International Journal of Business, Entrepreneurship and SMEs*, vol. 3, 2021.

[9] S. Balasescu, N. A. Neacsu, C. E. Anton, and M. Balasesu, "Study on e-commerce in the clothing industry in romania," *TEXTEH Proceedings*, vol. 2019, 2019.

[10] A. S. B. Ramirez, B. A. S. Diestra, and M. A. C. Lengua, "Implementation of a virtual store to exponentiate the flow of product sales in a private company in the city of lima," 2021.

[11] E. S. Soegoto, N. A. Rizqi, I. S. Purwani, and Z. Zulkarnain, "Zionimeart app: Designing mobile application as a medium for selling anime t-shirts," pp. 61–70, 2022.

[12] V. Gomero-Fanny, A. R. Bengy, and L. Andrade-Arenas, "Prototype of web system for organizations dedicated to e-commerce under the scrum methodology," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2021.0120152

[13] A. Tupia-Astoray and L. Andrade-Arenas, "Implementation of an e-commerce system for the automation and improvement of commercial management at a business level," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2021.0120177

[14] M. Lazo-Amado, L. Cueva-Ruiz, and L. Andrade-Arenas, "e-business model to optimise sales through digital marketing in a peruvian company," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, 2021. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2021.0121184

[15] T. Tia, I. Nuryasin, and M. Maskur, "Model simulasi rational unified process (rup) pada pegembagan perangkat lunak," *Jurnal Repositor*, vol. 2, 2020.

[16] A. P. Wibawa, M. Ashar, and S. Patmanthara, "Transfer teknologi pembuatan curriculum vitae dan poster untuk siswa pondok pesantren al-munawwaroh," *Belantika Pendidikan*, vol. 4, 2021.

[17] N. I. Yusman, "Perancangan sistem informasi berbasis orientasi objek menggunakan star uml di cv niasa bandung," *AIMS: Jurnal Accounting Information System*, vol. 1, 2018.

[18] H. D. Kuncahya, "Implementasi kotlin pada aplikasi pengenalan pahlawan nasional dan revolusi indonesia berbasis android," *Journal of Chemical Information and Modeling*, 2020.

[19] N. Lacey, "Sqlite," *Python by Example*, 2019.

[20] T. Hagos, "Android studio," *Learn Android Studio 4*, 2020.

Comparative Analysis of Machine Learning Algorithms and Data Mining Techniques for Predicting the Existence of Heart Disease

Nourah Alotaibi[1], Mona Alzahrani[2]
Department of Information and Computer Science (ICS)
KFUPM
Dhahran, Saudi Arabia

***Abstract*—Heart diseases are considered one of the leading causes of death globally over the world. They are difficult to be predicted by a specialist physician as it is not an easy task which requires greater knowledge and expertise for prediction. With the variety of machine learning and deep learning algorithms, there exist many recent studies in the state of the art that have been done remarkable and practical works for predicting the presence of heart diseases. However, some of these works were affected by various drawbacks. Hence, this work aims to compare and analyze different classifiers, pre-processing, and dimensionality reduction techniques (feature selection and feature extraction) and study their effect on the prediction of heart diseases existence. Therefore, based on the resulting performance of several conducted experiments on the well-known Cleveland heart disease dataset, the findings of this study are: 1) the most significant subset of features to predict the existence of heart diseases are PES, EIA, CPT, MHR, THA, VCA, and OPK, 2) Naïve Bayes classifier gave the best performance prediction, and 3) Chi-squared feature selection was the data mining technique that reduced the number of features while maintained the same improved performance for predicting the presence of heart disease.**

***Keywords*—Heart disease; feature selection; feature extraction; dimensionality reduction; Chi-squared; Naïve Bayes; Cleveland dataset**

I. Introduction

Cardiovascular diseases (CVDs) [1] are when the heart and blood vessels affected by some diseases like coronary heart disease and heart failure disease. The statistics in Saudi Arabia that were collected over the past 40 years indicate that the deaths have increased from CVDs. Moreover, according to the World Health Organization (WHO)[1], 17.9 million deaths every year resulted of CVDs including different heart diseases such as cardiovascular disease, valvular heart disease, heart defects, heart infections or cardiomyopathy [2].

Correctly predicting a diagnosis, including predicting the Existence of Heart Disease (EHD), is essential to patient-centered care, equally in choosing healing plans and notifying patients as a basis for shared decision making [3]. In recent years, there have been plenty of studies on EHD prediction. However, EHD prediction research has passed through three different stages along with history. In 1979, two researchers [4] combined diverse results gained from examinations like stress electrocardiography and cardiotocography, and others into a diagnostic decision about the likelihood of getting a disease in a particular patient through Bayes' Theorem . While in 1998, the second stage started when Wilson et al. [5] established a new direction concerning heart diseases estimation by utilizing risk factor classes with the aid of logistic approaches and regression calculations. Nowadays, several researchers have developed various machine learning algorithms [3, 6–10] to predict the EHD on the publicly available datasets which are the focus of this work.

However, the machine learning-based studies were affected by various drawbacks. For example, using datasets without handling the imbalance classes [6, 10, 11], important features such as age were manually excluded from the experiments [6, 9, 10], no comparisons were made in terms of prediction methods [6] or the used dataset [6, 11, 12], various platforms were used for assessments [10], some important details were missing or not clear such as the number of selected features [7] or the number of samples per class [3, 8, 11]. Even more, surprisingly, no coloration exists between the selected features among these works [3, 6–10] even though some of them were using the same dataset. In addition, different machine learning algorithms were selected as the best classifiers in various works according to their experimented data mining approaches. To sum up, there is still a need to experimentally evaluate different classifiers with different data mining approaches to gain a final decision. Hence, inspired by the development of several machine learning-based models for improving the EHD prediction, this study contributes to the literature by providing a work that handles these drawbacks as follows:

- Study the effect of different balancing solutions such as oversampling and undersampling for EHD prediction to handle the imbalance classes issue that exist on Cleveland heart disease dataset.
- Explore the most significant subset of features for the EHD prediction by analyzing different data mining techniques.
- Investigate the best performed classifier for the EHD prediction by comparing different machine learning algorithms.
- Achieve the highest performance for EHD prediction compared to recent related works that experimented

[1] World Health Organization. Global Health Observatory: Cardiovascular Diseases-Country Statistics. Retrieved on March 11, 2022 from: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1.

on the well-known Cleveland heart disease dataset by training the best performed classifier with the most significant subset of features.

Consequently, the following set of research questions were explored:

- RQ1: What are the best-performing feature sets for the EHD prediction?
- RQ2: How can the prediction performance be improved using data mining?
- RQ3: What is the most appropriate classifier using the selected features for EHD prediction?

The remaining of this paper is organized as the following: Section II presents the related works which provide several EHD prediction machine learning-based models. Section III explains the proposed methodology. Section IV demonstrates the achieved outcomes and tackles the discussion in light of the findings. Finally, Section V concludes this study and discuses future works.

II. Literature Review

Several heart diagnosis studies in the state of the art [3, 6–14] have been done extraordinary works that contributed by providing different prediction approaches. These studies could be categorized based on the targeted prediction such as Heart Failure (HF) prediction [7], mortality or hospitalization prediction of the HF patient [3, 6], and EHD prediction [8–14]. In addition, they also could be categorized in terms of the investigated learning technique, whether it is supervised learning [3, 8, 11, 12], ensemble learning [6, 12], deep learning [7, 9] or even hybrid learning [10]. Table I summaries the recent related works [3, 6, 7, 10–12] in terms of their experimented datasets, pre-processing techniques, learning methodologies, performance evaluation and drawbacks.

For HF prediction, Maragatham et al. [7] took advantage of the availability of the intensive substantial historical information in Electronic Health Record (EHR) and related time stamped data, in which, the authors inspected whether the usage of deep learning would improve the model performance for early HF diagnosis. The tested data consists of 365,446 patients, where 4289 of them had HF. The examination of time stamped EHRs aided in identifying the relations between numerous diagnosis events and predicting when a patient is being examined for a disease. Medical concept vectors and Long Short-Term Memory (LSTM) network were used to determine the diagnosis events and HF prediction. The proposed model was trained using one-hot vectors and grouped code vectors. As an activation function, SiLU and tanh were used in the hidden layers, while in the output layer, Softmax was used. For weight optimization through the network, Bridgeout, a regularization technique was used. K-nearest neighbour (KNN) and SVM were implemented using Python Scikit-Learn version 0.16.1 while Theano 0.7 was used to implement LSTM network, multilayer perceptron (MLP), and Logistic Regression (LR) models. They conducted two different experiments according to the length of the prediction and observation windows; and the performance was compared to well-known supervised approaches such as LR, MLP, SVM and KNN. Specifically, the first experiment, when using 12-months and 6-months as observation and prediction windows respectively, gave an AUC of 0.797, and when using 18-months and 0-months gave 0.894 AUC.

While for mortality prediction of the HF patient, Adler et al. [6] developed MARKER-HF, a tool that computes a risk score between -1 and +1 to predict mortality of hospitalized and ambulatory HF patients as high risk or low risk mortality using ensemble learning. MARKER-HF is based on a machine learning model that is trained using AdaBoost which is a boosted decision tree algorithm that is implemented in the TMVA toolkit. They used eight features to train their models with data of 5822 patients taken from the A systems BIOlogy Study to TAilored Treatment in Chronic Heart Failure (BIOSTAT-CHF) project, University of California San Diego (UCSD) and San Francisco (UCSF) Medical Centers. MARKER-HF results showed its ability to predict mortality consistently in three different datasets. For mortality prediction with a 95% confidence interval for the area under the ROC curve, they achieved AUC=0.88 using UCSD, AUC=0.84 using UCSF, and AUC=0.81 using BIOSTAT-CHF.

Moreover, Angraal et al. [3] compared five different supervised learning classifiers not only for mortality prediction but also for hospitalization of HF outpatients with preserved ejection fraction (HFpEF) through three years of follow-up. They trained the following five methods: two LR, one with a forward selection and one with a lasso regularization for feature selection, gradient descent boosting, Random Forest (RF) and Support Victor machine (SVM). They used a total of 86 candidate features to train their models, including demographic, clinical, laboratory and electrocardiography data, and KCCQ scores obtained from the patients. These patients' data are taken from the Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) trail. The authors used 5-fold cross-validation to divide the learning set into five subsets, where 80% of them were used for training and 20% for testing. They experimentally proved that RF is the best classifier with 95% confidence interval achieved AUC=0.72, and Brier score=0.17 for mortality prediction; and AUC=0.76 and Brier score=0.19 for HF hospitalization prediction. Even more, they found that the best features to predict mortality are the body mass index, BUN levels, and KCCQ scores; where BUN levels, hemoglobin level (H) and KCCQ scores are the best features to predict HF hospitalization.

On the other hand, most of the heart diagnosis studies were focused on predicting the EHD [8–14]. Ananey-Obiri and Sarku [11] investigated the traditional supervised learning algorithms and data mining techniques for EHD prediction. They investigated the following three supervised algorithms: decision tree (DT), LR, Gaussian Naïve Bayes (NB). They experimented the well-known Cleveland heart disease datasets with all of its 13 features shown in Table II which are age, sex, Resting Blood pressure (RBP), Chest Pain Type (CPT), Serum Cholesterol (SCH), Fasting Blood Sugar (FBS), Maximum Heart Rate achieved (MHR), Resting Electrocardiographic Results (RES), Exercise Induced Angina (EIA), Peak Exercise Slope (PES), Old Peak (OPK), Thallium Scan (THA) and number of major Vessels Colored by Fluoroscopy (VCA). The tested dataset contained 287 observations out of 303 after the duplicated, missing values and outliers were removed as a pre-processing step. In addition, they used feature normalization as a feature scaling technique, and single value decomposition (SVD) as feature extraction to reduce the number of features from 13 to 4. The data was labeled as absent or present of heart diseases. The authors used 10-fold cross-validation to divide the learning set into ten subsets, where nine of them were used for training and one of them for testing. The reported results were 79.31% accuracy and 0.81 AUC for DT model, 76% accuracy and 0.87 AUC for Gaussian NB model, and 82.75% accuracy and 0.86 AUC for LR model. Moreover, in the work conducted by Reddy et al. [12], ten different supervised and ensemble learning techniques were tested for EHD prediction. These techniques include NB, LR, Sequential minimal optimization (SMO), bootstrap aggregation, AdaBoost, JRip, RF, and KNN. Cleveland dataset was tested with 303 samples and 13 pre-mentioned features. Three different feature selection methods were used to enhance the performance, which are chi-squared, BestFirst search method and ReliefF. 11 out of 13 features were selected with the best performance result. As an evaluation scheme, 10-fold cross-validation was applied. The best result was 85.15% accuracy which was obtained using the SMO classifier with Chi-Squared feature selection technique. Even more, a hybrid learning approach was adapted by Abdeldjouad et al. [10], in which a hybrid approach of various machine learning methods was proposed to predict EHD. These methods include AdaBoostM1, LR, Fuzzy Unordered Rule Induction (FURIA), Multi-Objective Evolutionary Fuzzy Classifier (MOEFC), Fuzzy Hybrid Genetic Based Machine Learning (FH-GBML), and Genetic Fuzzy System-LogitBoost (GFS-LB). They also experimented Cleveland database for training and assessing their methods using 10-fold cross-validation. Two models were built in which the first model used AdaBoostM1, LR, and MOEFC with 14 features, and reduced to 12 by removing the personal information (e.g. age and sex) with the wrapper feature selection method. While the second model used FURIA, GFS-LB, and FH-GBML and reduced the features to 6 with Principal Component Analysis (PCA) as a dimensionality reduction technique. The first model was selected using the majority voting as the final best-performed model with 80.20% accuracy. For conducting this work, the Keel tool was used for feature selection, while the Weka tool was used for feature extraction.

However, the EHD prediction studies [8–14] have some drawbacks. For example, using datasets without handling the imbalance classes [6, 10, 11], important features such as age, were manually excluded [6, 9, 10], no comparisons were made in terms of prediction methods [6] or the used dataset [6, 11, 12], various platforms were used for assessments [10], some important details were missing or not clear such as the number of selected features [7] or the number of samples per class [3, 8, 11]. Even more, surprisingly, no coloration exists between the selected features among these works [3, 6–10] even though some of them were using the same dataset. In other words, the significant factors that cause variance in recent proposed works' performance are still not fully investigated. To sum up, there is still a need to experimentally evaluate different classifiers with different data mining approaches to gain a final decision. Hence, this study's primary goal is to compare and analyze different classifiers and data mining techniques (feature selection and feature extraction) and their effect on improving the EHD prediction.

III. Methodology

This work handled the previously mentioned drawbacks, in which it explored numerous solutions including the usage of different sampling techniques to overcome imbalanced datasets issue represented in [6, 10] with different dimensionality reduction techniques. Moreover, all the important features were taken into consideration which were excluded manually by [6, 9, 10]. Furthermore, some of the previous works [3, 6, 7] used different sets of performance metrics which make the comparison quite difficult, so the proposed approach was assessed by considering a full set of well-known performance metrics with clarifies details regarding the selected features and the number of samples per class to tackle what was missing in [3, 7, 8]. In addition, different machine learning algorithms were selected as the best classifiers in various works according to their experimented data mining approaches. To summarize, there is still a need to experimentally evaluate different classifiers with different data mining approaches to gain a final decision.

Consequently, the methodology shown in Fig. 1 which consists of seven main stages, was designed to conduct this study to ensure the proposed solutions. The first stage is *data collection* where the Cleveland heart disease dataset (Clev) [15] was selected in Section III-A to be investigated in this work. While in the second stage, a *data pre-processing* is performed in Section III-B using some data balancing techniques to generate two more balanced versions of the dataset. In Section III-C, *dimensionality reduction* was done as the third stage using some data mining techniques such as *feature selection*, and *feature extraction* to reduce the number of features, improve the performance and avoid overfitting. In Section III-D the fifth stage, which is *evaluation scheme preparing* was detailed. Then, seven well-known *classification algorithms* are selected in Section III-E for the purpose of comparison. Lastly, in the seventh stage, the conducted *comparative experiments* were designed in Section III-F.

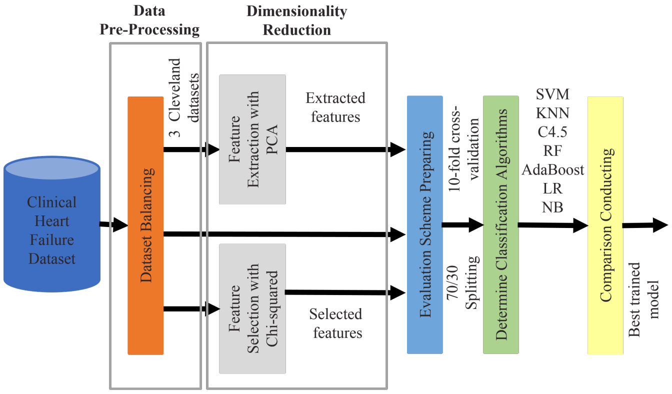


Fig. 1. The Proposed Methodology for EHD Prediction Comparison.

A. Dataset Collection

In this study, the well-known Clev heart disease dataset [15] was experimented since it is the most investigated dataset in this field by related works [8–11, 13, 14]. It is publicly available on an online machine learning and data mining repository of the University of California, Irvine (UCI). It contains

TABLE I. Summary of Various HF Prediction Techniques

	Reference	**Adler et al. [6]**	**Maragatham et al. [7]**	**Ananey-Obiri and Sarku [11]**	**Angraal et al. [3]**	**Reddy et al. [12]**	**Abdeldjouad et al. [10]**
Description	**Year**	2020	2019	2020	2020	2021	2020
	Main Goal	Mortality prediction of HF patients	HF prediction	EHD prediction	Mortality and hospitalization prediction of HF patients	EHD prediction	EHD prediction
Dataset	**Dataset Name**	UCSD	An arbitrary	Cleveland	TOPCAT	Cleveland	Cleveland
	# Samples	5822	4289	287	1,76	303	296
	Evaluation Scheme	50%|50% training|testing	6-fold cross-validation	10-fold cross-validation	5-fold cross-validation	10-fold cross-validation	10-fold cross-validation
Data Mining and Pre-Processing	**Technique**	- Exclusion of patients who had missing data, older than 80 years, with CIED device, died within 7 days of initial encounter, or had obvious medical record errors	Not clear	- Exclusion of patients who had missing or duplicated data - Feature scaling using normalization - Feature extraction using SVD	- Exclusion of features with 50% missing data - Feature selection using forward selection and a lasso regularization	- Feature selection using chi-squared, BestFirst search method and ReliefF	- Exclusion of patients who had missing data - Exclusion of personal features - Feature selection using a wrapper method - Feature extraction using PCA
	#Features	8	Not clear	4 out of 13	86	11 out of 13	13
	Features Names	Cr, RBP, H,BUN, platelets, WBC, RDW and albumin	Health info, tobacco usage, demographics and liquor consumption and lab test	Age, sex, RBP, CPT, SCH, FBS, MHR, RES, EIA, PES, OPK, THA and VCA	Demographic, clinical, laboratory and electrocardiography, and KCCQ scores features	Age, sex, RBP, CPT, SCH, FBS, MHR, RES, EIA, PES, OPK, THA and VCA	RBP, CPT, SCH, FBS, MHR, RES, EIA, PES, OPK, THA and VCA
Methodology	**Learning Category**	Ensemble learning	Deep learning	Supervised learning	Supervised and ensemble learning	Supervised and ensemble learning	Hybrid learning
	Prediction Methods	AdaBoost	LSTM	LR	LR with a forward selection, LR with a lasso regularization, RF, gradient descent boosting and SVM	SMO	A new hybrid approach of LR, AdaBoostM1, MOEFC, FURIA, GFS-LB and FH-GBML
Performance Evaluation	**Metrics**	ROC charts, and AUC	ROC charts, and AUC	CM, ACC, P, SE, F-score, ROC charts, and AUC	AUC, and Brier scores	ACC, MAE, SE, fallout, P, F-Score, SP, and ROC area	SE, SP, ACC, ER
	Results	AUC= 0.88	AUC= 0.894	ACC=82.75%, and AUC=0.86	Mortality prediction (AUC= 0.7, Brier score= 0.17) HF hospitalization prediction (AUC= 0.76, Brier score= 0.19)	ACC= 86.468%	ACC= 80.20%
Drawbacks	**In terms of** datasets, features, platforms, algorithms, and comparisons	- Imbalanced datasets - Exclusion of elderly patients above 80 years - No comparisons are made in terms of prediction methods	- Some important details are missing or not clear	- No. samples per class were not mentioned - Imbalanced datasets	- No. samples per class were not mentioned - No comparisons are made in terms of datasets - Dataset with missing data	- No comparisons are made in terms of the dataset	- Personal information is excluded manually (e.g age, sex) - Using various platforms - No comparisons are made in terms of the dataset - Imbalanced datasets

The evaluation metrics are P: Precision, CM: Confusion matrix, AUC: Area Under Curve, ROC: Receiver Operating Characteristic curve, ACC: Accuracy, SP: Specitivity, SE: Sensitivity, ER: Error Rate, MAE: Mean Absolute Error.
The features are RBP: Resting Blood pressure, CPT: Chest Pain Type, FBS: Fasting Blood Sugar, SCH: Serum Cholesterol, MHR: Maximum Heart Rate achieved, RES: Resting Electrocardiographic Results, EIA: Exercise Induced Angina, PES: Peak Exercise Slope, OPK: Old Peak, THA: Thallium Scan, VCA: Number of Major Vessels Colored by Fluoroscopy.

303 medical records of 165 patients with heart diseases and 138 are healthy. Moreover, it has 74 features, but 13 common features have been studied in the state of the art. Table II lists these features, their types and descriptions. The Clev originally contained five categorical classes (0, 1, 2, 3 and 4) where 0 refers to the absence of heart disease while the other four classes (1, 2, 3 and 4) refer to the presence of different heart diseases. However, most of the works [8–14] that used this dataset transferred these five categorical classes to a binary class for the purpose of distinguishing simplicity. The “class" field denotes the existence of heart disease in the patient. In which 0 means absence of heart disease (normal) and 1 means existence of heart disease.

B. Dataset Balancing

In the medical field, according to [10], the diagnosis of diseases is easier and quicker if data is balanced. The used Clev dataset contains 303 observations which present quite balanced positive and negative samples, with 165 and 138 observations respectively. But, machine learning techniques are very sensitive and the created prediction models are usually biased towards the larger class. However, the previous works [6, 10, 11] ignored this fact. Hence, to create an unbiased prediction model, reduce the gap between the two classes, and ensure equivalent balancing, dataset pre-processing is needed [9]. In this study, the class-imbalance problem in the original Cleveland dataset was solved by using oversampling and un-

TABLE II. Detailed Description of Cleveland Dataset's Features

#	Feature Name	Shortcut	Feature Type	Description
F1	Age	Age	Continuous	Age of the patient [years]
F2	Sex	Sex	Discrete	Sex of the patient [M: Male, F: Female]
F3	ChestPainType	CPT	Discrete	Chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
F4	RestingBP	RBP	Continuous	Blood pressure at rest [mm Hg]
F5	Cholesterol	SCH	Continuous	Serum cholesterol [mm/dl]
F6	FastingBS	FBS	Discrete	Fasting blood sugar [1: if FastingBS >120 mg/dl, 0: otherwise]
F7	RestingECG	RES	Discrete	Resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of >0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]
F8	MaxHR	MHR	Continuous	Maximum heart rate achieved [Numeric value between 60 and 202]
F9	ExerciseAngina	EIA	Discrete	Exercise-induced angina [Y: Yes, N: No]
F10	Oldpeak	OPK	Continuous	Oldpeak = ST depression induced by exercise relative to rest [Numeric value measured in depression]
F11	ST_Slope	PES	Discrete	The slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]
F12	MajorVessels	VCA	Continuous	The number of major vessels colored by flourosopy [0-3]
F13	ThalliumScan	THA	Discrete	Type of defect [3 = normal; 6 = fixed defect; 7 = reversible defect]

dersampling techniques to create two balanced versions of the dataset. The first version is (*Oversampled Cleveland*) which was generated using the Synthetic Minority Oversampling Technique (SMOTE) [16] which is popularly used in the medical field to deal with class imbalanced data. SMOTE adds more samples to the smaller class observations by generating random synthetic ones from its nearest neighbors using the Euclidean distance [17]. SMOTE increased the original Cleveland dataset from 303 to 330 observations (165 for each class). While the second version is (*Undersampled Cleveland*) which was simply generated by randomly reducing the number of larger class observations. Undersampling reduced the original Cleveland dataset from 303 to 276 observations (138 for each class).

C. Dimensionality Reduction

It decreases the number of input features (dimensions) of the original problem using specific techniques to improve the learning performance. These techniques are categorized as *feature selection* and *feature extraction*. The key difference between them is that the feature selection selects a subset of features from the original features, while the feature extraction uses the original features to create a new set of features [18]. In this study, one technique from each of these categories was experimented in the following illustration:

1) Feature Selection: it is the process of selecting a group of relevant features from the original features according to specific criteria [10]. Some of its main goals are: 1) reduce the algorithm's computational time, 2) identify the relevant features, 3) improve the prediction performance, and 4) avoid the overfitting by limiting the number of selected features; because the overfitting could affect the model to loss its robustness when the model is used to test new unseen data [6, 19]. In this study, the Chi-squared [20] was used as a feature selection technique to determine the most relevant features. Chi-squared was selected among other feature selection techniques since it improves most of the classifiers' performances and achieves remarkable results in this field [12, 14, 20]. This study starts with the 13 most common features and end up with only seven features after applying Chi-squared. The seven selected features are PES, EIA, CPT, MHR, THA, VCA, and OPK.

2) Feature Extraction: as one of the dimensionality reduction methods, it reduces the number of dataset's features in which the reduced features are represented by a set of new features [10]. The main goal of this technique is to use fewer features, which results in a simpler model that may have better performance with new unseen data. In this study, the principal component analysis (PCA) was used. The reason behind selecting PCA is because it is considered as one of the most famous dimensionality reduction and feature (components) extraction techniques for the medical applications [10]. PCA works by creating novel factors that have the best valuable information by capturing the highest variance of these features [21]. Using PCA, 2, 8 and 8 new sets of features were extracted for the original Clev, Oversampled Clev, and Undersampled Clev, respectively.

D. Evaluation Scheme Preparing

All the three Cleveland versions were evaluated by two schemes which are 10-fold cross-validation and 70%|30% for training|testing data splitting.

E. Classification Algorithms Selection

Machine learning and classification techniques are a group of computational models that can be used to solve many kinds of problems easily. Various applications of computational intelligence exist in the pathology and medicine field [22, 23]. This work, compared the following seven classifiers: SVM [24], KNN [25], C4.5 [26], RF [27], AdaBoost [28], NB [29] and LR [30]. SVM is a supervised learning technique that demonstrates superb performance in the medical field [31]. It depends on kernel functions that transfer all instances to a upper dimensional space intending to find a linear decision boundary for data partitioning [24]. KNN is a simple but effective method for classification [25]. C4.5 decision tree was selected due to its low complexity in implementation and excellent explanation [26]. In addition, a decision tree was investigated as the main classifier in several EHD prediction research. RF decision tree [27], is one of the popular techniques for pattern recognition which has been efficiently applied as a strong and widespread tool for predicting and classifying medical data. NB [29] is based on Bayes' Theorem with an assumption of independence among predictors.

In this study, SVM was implemented via the LibSVM library using nu-SVC as SVM type and linear kernel [24]. KNN was implemented using IBk library where the number of neighbors K to inspect equals 1. C4.5 decision tree algorithm was implemented using the J48 classifier (C4.5 release 8 implemented with Java) [26]. AdaBoost [28] is a short term of Adaptive Boosting and was implemented using AdaBoostM1.

F. Comparison Conducting and Performance Measurements

The comparison experiments were conducted based on the three versions of the Clev dataset (original, oversampled, and undersampled Clev), each with three copies: 1) without dimensionality reduction, 1) after applying Chi-squared, 3) after applying PCA; which makes them a total of nine datasets to be experimented. Each dataset was used to build training models using the seven classifiers which are SVM [24], KNN [25], C4.5 [26], RF [27], AdaBoost [28], NB [29] and LR [30]. So, a total of 63 models (9 datasets * 7 classifiers) were experimented. Each model was tested using two evaluation schemes; 10-fold cross-validation and 70%|30% splitting. Hence, the entire experiment ended up with a total of 126 trails (63 training models * 2 evaluation schemes).

The building training models were evaluated via six measures, which are Matthews correlation coefficient (MCC) [8], accuracy (ACC), recall/sensitivity (SE), precision, F-Score (FM) [11], specificity (SP), error rate (ER) [10], and area under the curve (AUC) [7]. MCC is a measure that is frequently utilized for assessing the quality of binary classification. It ranged from -1 to 1, in which 1 indicates an excellent prediction, 0 means the classification is no better than a random prediction, and -1 indicates full disagreement between prediction and observation. The Receiver Operating Characteristic (ROC) chart is used for additional investigation, where it consists of two rates, the true positive rate (TPR) versus the false positive rate (FPR) for various thresholds. The best ROC is the chart with more area under the curve (AUC). AUC equal to 1 presents the ideal ROC which means that the model can perform with 100% sensitivity and 100% specificity [8]. ACC refers to the fraction of accurately classified samples. SE is the fraction of correctly classified heart disease patients. It is also known as True Positive Rate (TPR) or recall, while SP is the correctly classified healthy subjects. These measures are formulated as the following[2]:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall/\ Sensitivity\ (SE) = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity\ (SP) = \frac{TN}{TN + FP} \tag{3}$$

$$F\text{-}Score = \frac{2TP}{2TP + FP + FN} \tag{4}$$

$$Accuracy\ (ACC) = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Error\ Rate\ (ER) = \frac{FP + FN}{P + N} \tag{6}$$

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP) * (TP + FN)(TN + FP)(TN + FN)}} \tag{7}$$

All the implemented experiments including oversampling, feature extraction, feature selection, and performance evaluation were done using various libraries in WEKA version 3.8.5 [32]. The oversampling was implemented using the *SMOTE* technique that was proposed by [16] under the supervised filters where the nearest neighbors parameter was set to 5. While feature extraction was implemented using *Principal-Components* as attribute evaluator and *Ranker* as the search method. On the other hand, feature selection was implemented using *ChiSquaredAttributeEval* as an attribute evaluator and *Ranker* as a search method as well.

IV. Results and Discussion

Tables III and IV summarize the splitting and cross-validation results respectively. These results only present the best performed classifiers of the nine datasets that have been obtained from 126 trials. While Fig. 2 and 3 present the splitting and cross-validation results in terms of accuracy.

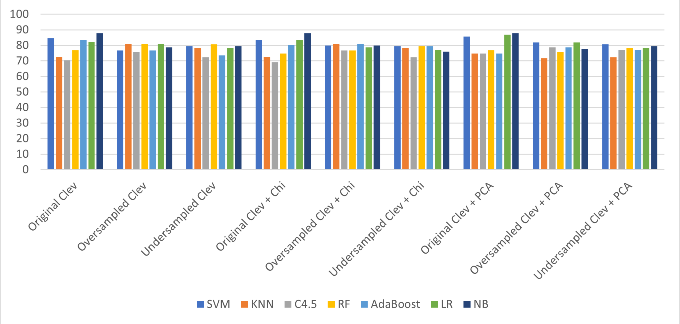


Fig. 2. The Accuracy of the Conducted Trails using 70%|30% Splitting.

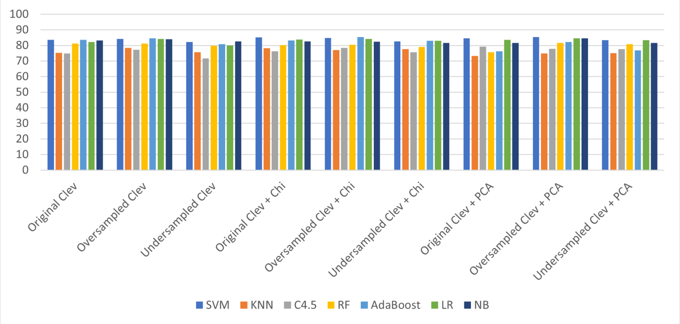


Fig. 3. The Accuracy of the Conducted Trails using 10-Fold Cross-Validation.

For *Original Clev dataset*, it can be noticed from the obtained accuracy in Fig. 2 of the splitting scheme that there are no significant differences between all the three versions of it (without processing, with Chi-squared, and with PCA) even though they contain different number of features. In fact, NB is the best performing classifier with ACC=87.91% and AUC=0.93, which is also summarized in Table III. C4.5 is the worst classifier, even when the number of features are changed. On the other hand, in terms of accuracy, when the cross-validation was used as shown in Fig. 3, SVM and AdaBoost were the best performing classifiers using all the 13 features with the same 83.5% accuracy. But SVM

[2] Confusion Matrix: https://en.wikipedia.org/wiki/Confusionmatrix

TABLE III. The Summarized Results using the 70%|30% Splitting as Evaluation Scheme

Exp #	Dataset	Data Mining Tech.	# of Features	Best Classifier	SP	SE/Recall	Precision	ACC	ER	FM	MCC	AUC
1	Original Clev	Without	13	NB	0.86	0.90	0.84	**87.91**	0.12	0.87	0.76	0.93
2		Chi-squared	7	NB	0.86	0.90	0.84	**87.91**	0.12	0.87	0.76	0.93
3		PCA	2	NB	0.86	0.90	0.84	**87.91**	0.12	0.87	0.76	0.93
4	Oversampled Clev	Without	13	LR RF	0.73	0.90	0.75	80.81	0.19	0.82	0.63	0.89
5		Chi-squared	7	AdaBoost	0.76	0.85	0.77	80.81	0.19	0.81	0.62	0.88
6		PCA	8	LR	0.73	0.92	0.76	**81.82**	0.18	0.83	0.65	0.88
7	Undersampled Clev	Without	13	RF	0.71	0.90	0.76	**80.72**	0.19	0.82	0.63	0.85
8		Chi-squared	7	RF	0.76	0.83	0.77	79.52	0.20	0.80	0.59	0.84
9		PCA	8	SVM	0.79	0.83	0.79	**80.72**	0.19	0.81	0.62	0.81

TABLE IV. The Summarized Results using the 10-Fold Cross-Validation as Evaluation Scheme

Exp. #	Dataset	Data Mining Tech.	# of Features	Best Classifier	SP	SE/Recall	Precision	ACC	ER	FM	MCC	AUC
1	Original Clev	Without	13	AdaBoost	0.78	0.88	0.83	83.50	0.17	0.85	0.67	0.88
2		Chi-squared	7	SVM	0.77	0.92	0.83	**85.15**	0.15	0.87	0.70	0.84
3		PCA	2	SVM	0.83	0.86	0.86	84.49	0.16	0.86	0.69	0.84
4	Oversampled Clev	Without	13	AdaBoost	0.85	0.84	0.85	84.55	0.15	0.84	0.69	0.85
5		Chi-squared	7	AdaBoost	0.85	0.86	0.85	**85.45**	0.15	0.86	0.71	0.92
6		PCA	8	SVM	0.85	0.85	0.85	**85.45**	0.15	0.85	0.71	0.85
7	Undersampled Clev	Without	13	NB	0.80	0.85	0.81	82.61	0.17	0.83	0.65	0.89
8		Chi-squared	7	AdaBoost	0.81	0.85	0.82	82.97	0.17	0.83	0.66	0.89
9		PCA	8	LR	0.82	0.85	0.82	**83.33**	0.17	0.84	0.67	0.89

TABLE V. The Most Recent Works that Studied the Cleveland Dataset

Ref.	Year	# of Features	Applied Technique	Classification Methods	Evaluation Scheme	ACC	AUC	MCC
[13]	2020	14	Data pre-processing	Artificial Neural Network (ANN)*, LR, SVM, KNN, NB and RF	10-fold cross-validation	85.86	-	-
[14]	2019	9	Feature selection	Majority voting of the weak classifiers	Splitting	85.48	-	-
[11]	2020	4	Data pre-processing, feature scaling, feature extraction, and outlier detection	Gaussian NB*, DT, and LR	10-fold cross-validation	82.75	0.87	-
[33]	2020	10	Data pre-processing, feature scaling, and data reduction	RF*, SVM, KNN, DT and LR	Splitting	85.71	0.87	-
Ours	2021	7	Data pre-processing, feature selection, feature scaling, and feature extraction	NB*, SVM, KNN, C4.5, RF, AdaBoost and LR	Splitting	**87.91**	**0.93**	**0.80**

* mean the classifier that gave the best performance in terms of accuracy.

outperformed the others when the features are reduced to 7 and 2 using Chi-squared (ACC=85.15% and AUC=0.84) and PCA (ACC=84.49% AUC=0.84).

For *Oversampled Clev dataset*, when the SMOTE technique is used to balance the data, the accuracy of the splitting scheme in Fig. 2 was significantly changed when the processing technique was changed (# of features) and with the evaluation scheme being changed. But the LR was the best classifier with ACC=81.82% when only the 8 features obtained from PCA were used. C4.5 and KNN gave the worst accuracy. However, in terms of AUCs, the RF and LR always outperformed other classifiers, even when the number of features being changed by 0.88 AUC for both classifiers. In addition, in terms of accuracy using cross-validation shown in Fig. 3, C4.5 and KNN again gave the worst accuracy. Where AdaBoost dominates the other classifiers by ACC=85.45%, when the original 8 or 7 features from Chi-squared and PCA were used, respectively. Moreover, in terms of AUCs, the best classifiers vary when the number of features being changed, but the worse classifiers were always C4.5 and KNN.

For *Undersampled Clev dataset*, Fig. 2 and 3, show that when the original Clev observations were reduced aiming for balancing, the obtained results were decreased compared with the above versions of the dataset. However, with the splitting scheme, it achieved 80.72% accuracy when using the 13 features with RF, and also when using the 8 features from PCA with SVM. Where with cross-validation scheme, it achieved 82.97% accuracy when using the 7 features from Chi-squared with AdaBoost. Furthermore, the LR and NB are the best classifiers in terms of AUCs, whatever the used features. Even with the splitting scheme, it achieved AUC=0.85 when using the 13 features with RF. Yet, with the cross-validation scheme, it achieved AUC=0.85 with whatever the used features.

According to the obtained results from the above experiments, the research questions were answered as follows: 1) seven features, which are PES, EIA, CPT, MHR, THA, VCA, and OPK, are the best performing features for predicting EHD, which were obtained from the Chi-squared feature selection technique; 2) it is noticeable that the best prediction performance ACC=87.91% and AUC=0.93 can be obtained whither the original features, the selected features, or the extracted features were used. Hence, the data mining techniques did not improve the prediction performance in terms of accuracy, however, they improved it in terms of reducing the number of features which lead to more computational efficiency; and 3) NB proves that it is the most appropriate classifier to be used with whatever feature sets to predict the EHD.

Furthermore, the availability of the Clev dataset allows many researchers to test their prediction models. For that reason, this work was compared to the recent related studies that used Clev dataset [11, 13, 14, 33]. Table V summarises their methods and obtained results along with this work. It is noticeable that these studies lack some important performance metrics such as MCC which play an important role in reporting balance or imbalanced data. Moreover, the effect of using different applied techniques such as data pre-processing, feature selection, and feature extraction techniques were also compared. This comparison proved that this work outperforms those studies by achieving higher performance with a margin of 2.20%.

V. Conclusion and Future Work

EHD prediction is a field where researchers propose new techniques that hopefully can facilitate the diagnosis of the existence of heart diseases and enhance the decision-making operations of physicians. In this work, an experimental comparison was conducted between seven famous classifiers, which are SVM, KNN, C4.5, RF, AdaBoost, NB, and LR with different data mining techniques including feature extraction, and feature selection. This work utilized a famous heart disease dataset called Cleveland, aiming to undergo a deeper investigation of the effective techniques that could improve EHD prediction. A methodology of seven basic stages was proposed to conduct this study, including data collection, data pre-processing, and balancing techniques (oversampling and undersampling). Dimensionality reduction such as Chi-squared feature selection and PCA feature extraction techniques were also investigated. The main concern of this paper is not only enhancing the accuracy of weak classifiers but also investigating the famous Clev dataset closely and studying the influence of different pre-processing techniques, in addition, to determining the number of best features that work better with Clev. However, this work, like other Cleveland dataset-based works [11, 13, 14, 33], suffers from the limited number of observations that could be handled in the future works by merging it with another EHD prediction dataset. Moreover, this study could be extended by exploring more data pre-processing techniques such as outlier detection, applying deep learning techniques, and tuning the hyperparameters.

Acknowledgment

We gratefully thank Dr. Imane Boudellioua for her valuable suggestions and guidance, sharing her encouragement and useful feedback in completing this work.

References

[1] N. M. Aljefree, I. M. Shatwan, and N. M. Almoraie, "Association between nutrients intake and coronary heart disease among adults in saudi arabia: A case-control study," *PROGRESS IN NUTRITION*, vol. 23, no. 3, 2021.

[2] Heart disease. Accessed 19-October-2021. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc-20353118

[3] S. Angraal, B. J. Mortazavi, A. Gupta, R. Khera, T. Ahmad, N. R. Desai, D. L. Jacoby, F. A. Masoudi, J. A. Spertus, and H. M. Krumholz, "Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction," *JACC: Heart Failure*, vol. 8, no. 1, pp. 12–21, 2020.

[4] G. A. Diamond and J. S. Forrester, "Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease," *New England Journal of Medicine*, vol. 300, no. 24, pp. 1350–1358, 1979.

[5] P. W. Wilson, R. B. D'Agostino, D. Levy, A. M. Belanger, H. Silbershatz, and W. B. Kannel, "Prediction of coronary heart disease using risk factor categories," *Circulation*, vol. 97, no. 18, pp. 1837–1847, 1998.

[6] E. D. Adler, A. A. Voors, L. Klein, F. Macheret, O. O. Braun, M. A. Urey, W. Zhu, I. Sama, M. Tadel, C. Campagnari *et al.*, "Improving risk prediction in heart failure using machine learning," *European journal of heart failure*, vol. 22, no. 1, pp. 139–147, 2020.

[7] G. Maragatham and S. Devi, "Lstm model for prediction of heart failure in big data," vol. 43, no. 5, 2019.

[8] L. Ali, A. Niamat, J. A. Khan, N. A. Golilarz, X. Xingzhong, A. Noor, R. Nour, and S. A. C. Bukhari, "An optimized stacked support vector machines based expert system for the effective prediction of heart failure," *IEEE Access*, vol. 7, pp. 54 007–54 014, 2019.

[9] L. Ali and S. Bukhari, "An approach based on mutually informed neural networks to optimize the generalization capabilities of decision support systems developed for heart failure prediction," *Irbm*, 2020.

[10] F. Z. Abdeldjouad, M. Brahami, and N. Matta, "A hybrid approach for heart disease diagnosis and prediction using machine learning techniques," in *International conference on smart homes and health telematics*. Springer, 2020, pp. 299–306.

[11] D. Ananey-Obiri and E. Sarku, "Predicting the presence of heart diseases using comparative data mining and machine learning algorithms," *International Journal of Computer Applications*, vol. 176, pp. 17–21, 2020.

[12] K. V. V. Reddy, I. Elamvazuthi, A. A. Aziz, S. Paramasivam, H. N. Chua, and S. Pranavanand, "Heart disease risk prediction using machine learning classifiers with attribute evaluators," *Applied Sciences*, vol. 11, no. 18, p. 8352, 2021.

[13] I. Tougui, A. Jilbab, and J. El Mhamdi, "Heart disease classification using data mining tools and machine learning techniques," *Health and Technology*, vol. 10, pp. 1137–1144, 2020.

[14] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," *Informatics in Medicine Unlocked*, vol. 16, p. 100203, 2019.

[15] M. Lichman. (2013) UCI machine learning repository. [Online]. Available: http://archive.ics.uci.edu/ml

[16] 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.

[17] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Improving the prediction of heart failure patients' survival using smote and effective data mining techniques," *IEEE Access*, vol. 9, pp. 39 707–39 716, 2021.

[18] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Computers in biology and medicine*, vol. 112, p. 103375, 2019.

[19] S. E. Awan, M. Bennamoun, F. Sohel, F. M. Sanfilippo, B. J. Chow, and G. Dwivedi, "Feature selection and transformation by machine learning reduce variable numbers and improve prediction for heart failure readmission or death," *PloS one*, vol. 14, no. 6, p. e0218760, 2019.

[20] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An automated diagnostic system for heart disease prediction based on x^2 statistical model and optimally configured deep neural network," *IEEE Access*, vol. 7, pp. 34 938–34 945, 2019.

[21] A. K. Gárate-Escamila, A. H. El Hassani, and E. Andrès, "Classification models for heart disease prediction using feature selection and pca," *Informatics in Medicine Unlocked*, vol. 19, p. 100330, 2020.

[22] A. H. Shahid and M. Singh, "Computational intelligence techniques for medical diagnosis and prognosis: Problems and current developments," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 3, pp. 638–672, 2019.

[23] H. R. Tizhoosh and L. Pantanowitz, "Artificial intelligence and digital pathology: challenges and opportunities," *Journal of pathology informatics*, vol. 9, 2018.

[24] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 1–27, 2011.

[25] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine learning*, vol. 6, no. 1, pp. 37–66, 1991.

[26] S. L. Salzberg, "C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993," 1994.

[27] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[28] Y. Freund, R. E. Schapire *et al.*, "Experiments with a new boosting algorithm," in *icml*, vol. 96. Citeseer, 1996, pp. 148–156.

[29] G. H. John and P. Langley, "Estimating continuous distributions in bayesian classifiers," *arXiv preprint arXiv:1302.4964*, 2013.

[30] S. Le Cessie and J. C. Van Houwelingen, "Ridge estimators in logistic regression," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 41, no. 1, pp. 191–201, 1992.

[31] J. Zhi, J. Sun, Z. Wang, and W. Ding, "Support vector machine classifier for prediction of the metastasis of colorectal cancer," *International journal of molecular medicine*, vol. 41, no. 3, pp. 1419–1426, 2018.

[32] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[33] N. K. Kumar, G. S. Sindhu, D. K. Prashanthi, and A. S. Sulthana, "Analysis and prediction of cardio vascular disease using machine learning classifiers," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2020, pp. 15–21.

Crop Field Monitoring and Disease Detection of Plants in Smart Agriculture using Internet of Things

Mr. G. Balram
Research Scholar, Department of CSE
KLEF
Vaddeswaram, A.P.

Dr. K. Kiran Kumar
Professor, Department of CSE
KLEF
Vaddeswaram, A.P.

***Abstract*—The Internet of Things can be defined as the network of physical objects that have sensors, software, and other technologies built into them in order to communicate and exchange data with other systems and devices over the internet. In intelligent agricultural advancements to increase the quality of agriculture, the Internet of Things (IoT) can be used. The manual monitoring of plant diseases is quite challenging. It demands enormous effort, expertise in the diseases of plants and the considerable time required for processing. The idea of automation in Smart Agriculture is implemented using the Internet of Things (IoT). They help monitor the plant leaf conditions, control water irrigation, gather images using installed IoT system which includes NodeMCU, cameras, soil moisture, temperature sensors and detect diseases in plants on the datasets collected from leaves. To detect plant diseases, image processing is applied. The detection of diseases comprises the acquisition of images, image pre-processing, segmenting an image, extracting and classifying features. In addition, the performance of two machine-learning techniques, such as a linear and polynomial kernel multi hidden extreme machine (MELM) and a support vector machine (SVM), has been studied. This paper discussed how plant diseases could be detected via images of their leaves. This analysis seeks to validate a proposed system for an appropriate solution to the IoT-based environmental surveillance, water irrigation system management and an efficient approach for leaf disease detection on plants. The proposed multi hidden layers extreme machine classification delivers good performance of 99.12% in the classification of leaf diseases in comparison to the Support Vector Machine classification, which gives 98%.**

Keywords*—*Image acquisition; segmentation; feature extraction; Internet of Things; plant disease

I. Introduction

Agriculture is the primary livelihood of Indian villagers; more has been done to boost yields and help farmers address issues like agriculture and plant disease. There have been significant mechanical and chemical advances. In this industry, however, there is little or less digitization. With the IoT growth, it is hoped that the farmers will make a digital agriculture system that will assist farmers in making educated decisions on their farms and help them address undesirable issues earlier. It will therefore avail to increase crop quality and will also benefit farmers. This will help—early disease detection, which in agriculture is a great challenge. IoT sensor nodes play a significant role in collecting real-time information play a significant role in collecting real-time information [1], [2]. These nodes can make the system highly practical by gathering data from crop fields in real-time to specify the agriculture system precisely. Presently to make agriculture system more practical with the help of machine learning techniques. All of them are more beneficial in different domains. Multiple applications are developed in precision agriculture to update the farmers about the condition of the crop [3]. The precision agricultural model is usually made up of three primary phases, as illustrated in Fig. 1. In the first phase, the availability of the IoT sensor nodes and their monitoring conditions can be obtained for crop. Later in the second phase, the data gathered from all the sensors are sent to the fog node with a wireless module to further process the data for higher levels where it can be monitored. Finally, the analytic methods are applied in the third phase of architecture to understand the status of the fields.

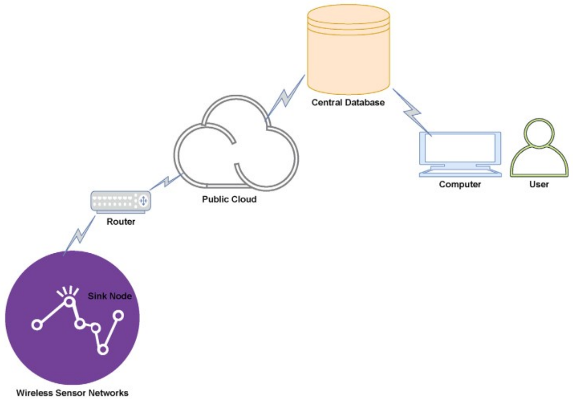


Fig. 1. Precision Agriculture Model.

An alert data is passed to the farmers to make them alert and take necessary precautions if any abnormal conditions are met, like dryness in the crop field, any leaves affected by diseases, with the help of deployed sensors. Therefore, there is continuous communication with the selector, which switches the dipping system on or off to allow the water for all the areas of the ground. Later, the farmers can use the required nitrogen, potassium, and phosphorus fertilizer to sustain plant growth and achieve crop yield. The response system is initiated by using analytics and actuators when any crucial situation is recognized (sensed/predicted). There are several IoT and WSN applications [4], some of which are discussed here. The applications for precision farming are shown in Fig. 2. The different implementation of greenhouses leads to specific significant planting problems and minor technical problems for the global conservation systems [5]—one of the primary issues employing temperature preservation. Pest insects and

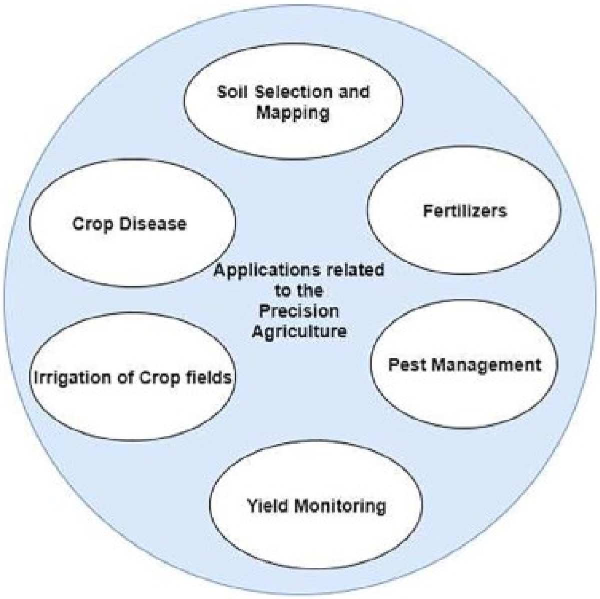

Soil Selection and Mapping
Crop Disease
Fertilizers
Applications related to the Precision Agriculture
Irrigation of Crop fields
Pest Management
Yield Monitoring

Fig. 2. Applications Related to Precision Agriculture.

flies have also damaged greenhouses because they are covered in a structure that generates turbulence in the covered sheet for any reptile or flies. This turbulence can eventually cause a reduction in plant health, and even their growth can be affected. Because of these issues, the greenhouse practice has to be evaluated and technology integrated to address these concerns. This paper emphasises applying the internet of things (IoT) to solve temperature monitoring and leaf dataset detection in the plant. It will allow agribusiness, farmers, horticultural companies and potential investors to transform the internet from an existing greenhouse into an innovative greenhouse.

II. Related Work

Various literature works on the cloud and edge in the health service have emerged. A software approach for the automatic detection and classification of plant leaf diseases was suggested and experimentally verified in [6]. The proposed system developed includes four fundamental steps. In the first step, most of the coloured pixels are detected. Those pixels are subsequently masked using the Otsu method depending on specific threshold values, and predominantly green pixels are masked. The second step is to remove pixels with red, green, and blue nulls and the pixels on the edges of the infected cluster. The experiment results show that the technique provided is a robust method for detecting diseases from plant leaf images. The efficiency of the proposed algorithm may successfully detect and classify the diseases studied with accuracy from 83In [7], the authors trained a deep convolutional neural network to detect 14 plant classes and 26 diseases from 54,306 images of unhealthy and healthy plant leaf taken under controlled conditions (or absence thereof). The trained model obtains a 99.35% accuracy with a stable test set, which shows the practicality of the technique. The model still reaches an accuracy of 31.4% when tested on a collection of images acquired from reputable web sources, i.e. under conditions different from those for training. Although this accuracy is substantially higher than that due to random selection (2.6%), additional training data are required to improve overall accuracy. Essentially, the technique of deep learning models in increasingly large and open-ended image data sets presents a holistic path toward that massive global diagnosis of smartphone-assisted plant disease. In [8], the authors integrated thermal and visible light image data with information on the depth. They custom-made a machine learning system for remotely detecting plants with Oidium neolycopersici powdery tomato fungus. The results make evident that, by integrating this information with the depth data, the detection accuracy of unhealthy plants has improved dramatically from the image data. Furthermore, it has been demonstrated that a new feature set may detect plants that were not initially injected with the fungus at the beginning of the experiment but became diseased through standard transmission. In [9], the authors studied several computer vision ways that resulted in the detection of multiple algorithms. Phase one was taken using a typical digital camera to gather images of banana leaves. Phase two involves using several extraction methods to collect essential data for phase three when images are healthy or unhealthy. Extremely randomized trees have done their best in detecting diseases with 0. 96 AUC in banana bacterial wilts and 0.91 in the case of banana black Sigatoka (BBS) among the seven classifiers used in this study. Finally, based on the area under curve (AUC) analysis, the performance of various classifiers was evaluated, and the optimum approach for the automatic diagnosis of these banana conditions was then selected. In [10], the nonexistence of farmers in a country where around 58 percent of the people participates in farming, the best crop for the land is decided on by traditional and non-scientific techniques. Farmers were sometimes unable to choose the suitable soil crops, season and geographical location. This leads to suicide, to an abandonment of agriculture and urban living conditions. The authors have presented a technique to help farmers select the crop to overcome this problem by considering all elements, such as the soil's season and the crop's location. Moreover, precise agriculture is practised in emerging countries with current agriculture techniques, focusing on site-specific crop management. In order to accurately separate the disease symptoms from the clutter background, manual selection of the input dataset was utilised at the same time. In [11], Textual features can be extracted in two different methods. One method for calculating the energy, inertia, entropy, and homogeneity of textual features is to utilise a grey level co-occurrence matrix. The Gabor transform can also be used. Some studies employ descriptors based on colour. Area, perimeter, centroid, and diameter were once taken into consideration for shapes. Due to its high accuracy and reduced dimensionality problem, Support Vector Machine (SVM) is a favoured classification method. In [12], the CNN is a neural network technology that is frequently used today to train or analyse visual data. Convolution's matrix format is intended to filter the images. The input layer, convo layer, fully connected layer pooling layer, drop-out layer to form CNN, and finally linked dataset classification layer are all used in the Convolution Neural Network for data training. In [13], to extract the distinctive elements from an image, segmentation is an approach that takes into account the texture and colour of the image. In [14],the adoption of deep learning, more specifically Convolutional Neural Network (CNN), as a substitute method for creating a model for classifying diseases. In [15], in order to extract local features, the convolution layer convolves the input image (or the output of the previous layer) using configurable weight filters, or kernel. CNN extracts shift-

and scale-invariant local features from the input by using various weight filters. The object's translation invariance is acquired by the pooling layer, which summarises the results of the preceding layer. In [16], color, texture, and shape are some of the features that were taken from the image. Information about boundaries, spots, and broken areas is contained in the colour feature. The percentage of the lesion and its type are also included in the shape attribute. Uniformity, contrast, probability, variance, and correlation are all aspects of texture. The database is split into two sets of photos for training and testing using the identified features. In [17], building a model that can be utilised by developers to build smartphone or online applications to detect tomato leaf diseases using convolutional neural networks in order to illustrate, categorise, and provide solutions for plant problems. In [18], the tool for classification is a neural network. Target data is provided to the neural network as a class vector, and seven extracted features—Contrast, Correlation, Energy, Homogeneity, Mean, Standard Deviation, and Variance—are given as input. In this case, the data was classified using a back propagation neural network. Following the network's training, it displays the performance plot, confusion matrix, and error histogram plot. In [19], a pixel's intensity in relation to its surrounding pixels throughout the entire image is measured as contrast. The basis for contrast in visual observations of the real world is how the colour and brightness of an object change in comparison to other objects in the frame of reference. A constant image has a contrast value of 0.

III. Materials and Methods

In this work, we concentrate on monitoring humidity, temperature, and water flow to know the environmental conditions of the crop field with the help of IoT sensors connected to the server. In addition, the work involves the early detection of plant leaf diseases by capturing images periodically with the proper deployment of cameras and applying image processing techniques. Our work mainly contributes two-fold: Initially, a standalone system for monitoring and irrigation, as shown in Fig. 3. Second, the analysis and detection for plant leaf disease as shown in Fig. 4. Different sensors such as soil moisture, temperature, and camera are used to detect diseases on a leaf. Data from sensors is collected and transmitted to NodeMCU via wired or wireless devices. Data are checked and matched to the ideal data values such as temperature, humidity, and soil moisture on the server-side (Cloud database). If there is a discrepancy in the threshold value, send the notification on the mobile or website to the farmer. In the webpage and farmer, the output of the sensors is generated.

A. NodeMCU

NodeMCU is an IoT platform open source. The ESP-12 module includes both firmware that operates on the ESP 8266 Wi-Fi SoC and hardware. A 32-bit RISC CPU is clocked at 80 MHz and supports up to 16MB external flash storage for a significant RAM supplement. Thanks to its small size and built-in Wi-Fi capability, the gadget is particularly ideal for IoT applications.

B. DHT11 Temperature and Humidity Sensor

For hotness measurement and environmental impact, a temperature sensor is utilized. Humidity is the amount of water vapour in the air that the hygrometer can measure. For this, we used the DHT11 sensor, which can measure the temperature and humidity data. This sensor has compactness and operates at low voltages of 2.2V, and draws a few mA currents, thus suited for all temperatures and humidity conditions at a long transmission distance of 20m.

C. Water Flow Sensor

The water flow sensor has a plastic valve design, a Hall Effect sensor and a water rotor. The motion of the rotor with a changing rate of flow alters as the water flows along with the rotor rolling unit for each revolution. The process of recording the output 5 to 6 pulses for every litre of liquid that flows for 60secs. In this current work, we used a water flow sensor with a 20mm diameter and water pressure of 1.75 MPa and a flow range of 1 25L per minute.

D. Soil Moisture Sensor

Understanding the soil moisture in our fields is critical for optimal irrigation scheduling for precision irrigation. That's why we use intelligent and precise soil moisture sensors, which may be used to water our plants just when they're dry, avoiding over-or under-watering. A standard soil moisture sensor consists of two parts: A fork-shaped probe with two bare conductors inserted into the soil or anywhere else that water content is to be recorded. As previously stated, it operates as a variable resistor, with resistance that varies depending on soil moisture. An electronic module is also included with the sensor, which couples the probe to the NodeMCU. The module generates an output voltage based on the probe's resistance and provides access via an Analog Output (AO) pin. The module comes with a built potentiometer for adjusting the digital output's sensitivity. Using a potentiometer, we can configure a threshold setpoint such that when the moisture level is above the predefined threshold, the module will output level '0' (LOW), otherwise '1'. (HIGH). This design is highly beneficial when we want to initiate an action when a given limit is met. We can, for instance, activate a relay to start pumping water when the moisture level in the soil reaches a certain level.

E. VGA Camera Module

The operation of the installed camera in the crop field permits regular access with a 24-7 clock visual display of the crop field and plant conditions. A compact image sensor, low voltage module 'OV7670 Camera' allows a single chip, a VGA camera and an image processor to complete. The sensor can deliver the full-frame, samples, and different resolutions of up to eight data bits with SCCB bus controls. With the VGA image that has been inserted, it is possible to obtain frame rates of up to 30 frames per second. It takes the leaf images with high resolution in a suitable format that may be extensively controlled and effectively measured. This procedure includes image processing features, including the white balance and correction in saturation levels interface. SCCB programming interface.

F. Disease Detection Framework

This subsection shows how the leaf diseases detection process can be done on leaf dataset images by applying image processing operations using Matlab application. All the gathered images are in RGB format.

G. Image Acquisition

The camera captures the images of the plant leaf. In RGB (Red, Green and Blue), this image has been generated. For the RGB image, a colour transformation structure is constructed, and a device-independent colour transformation is performed to the design of colour transformation.

H. Image Pre-Processing

Different preprocessing methods are proposed to eliminate noise from the image or other object removal. Cropping of the image, i.e. leaf image, to get the region of interest (ROI) of the image relevant. Image smoothing is done with filter smoothing. Enhance the effectiveness to increase the contrast. The RGB images are converted into grey images by the equation (1).

$$f(x) = 0.2989 * R + 0.5870 * G + 0.114. * B \tag{1}$$

The histogram equalization that distributes the image intensities is then applied to improve images of plant disease. Intensity levels are distributed by a cumulative distribution function [20].

I. Image Segmentation

Segmentation implies dividing the image into different parts or with certain similarities. Segmentation can be performed with several approaches such as Otsu, k-means, RGB-image conversion in HIS, etc.

J. Segmentation using Boundary and Spot Detection Algorithm

The image of RGB is transformed into a HIS segmentation model. Identifying the boundary and spots helps discover the part of the infected leaf [21]. In this current study, the 8-pixel connection is considered for boundary detection, and the algorithm for boundary detection is used.

K. Adaptive K-Means Clustering Algorithm (AK-Means)

This approach uses the maximum connected domain technique for the K-means segmentation method to determine K values. After many tests, the K value ranges typically from 2 to 5. The K-means strategy solves the problem of “Maximization of Expectations” to apply the assignment mechanism to the necessary information points for the closest cluster, as shown in equation (2). We can also observe from equation (4), that when the data points belong to a cluster of xi, they belong to other clusters of k. Likewise, μ_j term denotes explicitly cluster centre xi. Interestingly, for the sake of closed clusters, we used Adaptive K-Means for the clustering process of leaf images which are unhealthy for further processing.

$$C = \sum_{i=1}^{y} \sum_{j=1}^{z} w_{ij} x^i - \mu_j^2 \tag{2}$$

$$\frac{\partial C}{\partial w_{ij}} = sum_{i=1}^{y} \sum_{j=1}^{z} x^i - \mu_j^2 \tag{3}$$

$$w_{ij} = \begin{cases} {}^{1ifk=argmin_i x^i - \mu_j^2} \\ \\ {}_{0else} \end{cases} \tag{4}$$

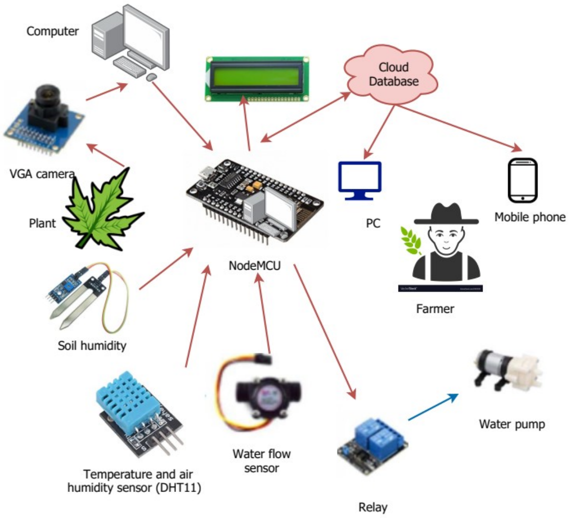


Fig. 3. Overview of Automated Monitoring and Irrigation System Hardware Design Architecture.

L. Feature Extraction

The extraction of features is a crucial element in object identification. The images were then removed by the Color Co-occurrence Method (CCM). The images were taken. The CCM consists of a matrix of pixel values distributed in the same way by the images at the provided offsets (grey scales or colours) in eq. (9). The features that can be employed to diagnose plant diseases are colour, texture, morphology, edges, etc. This method takes into consideration both colour and texture to create a unique image. The RGB is transformed to HSI translation for this purpose.

$$H = \begin{cases} {}^{\theta if B<G} \\ \\ {}_{360-\theta, B>G} \end{cases} \tag{5}$$

$$S = 1 - \frac{3}{(R+G+B)} \left[min\left(R, G, B\right) \right] \tag{6}$$

$$I = \frac{1}{3} \left(R + G + B\right) \tag{7}$$

The CCM approach employs SGDMs in which the CCM grey level is applied for doing the sampling process. The grey level is sampled in a way that is mainly associated with the other grey levels.

$$(p, q) = \sum_{x=1}^{n} \sum_{y=1}^{m} \begin{cases} {}^{ifl(x,y)=p,} \end{cases} 1, l\left(x + \Delta x, y + \Delta y = q\right)_{0else} \tag{8}$$

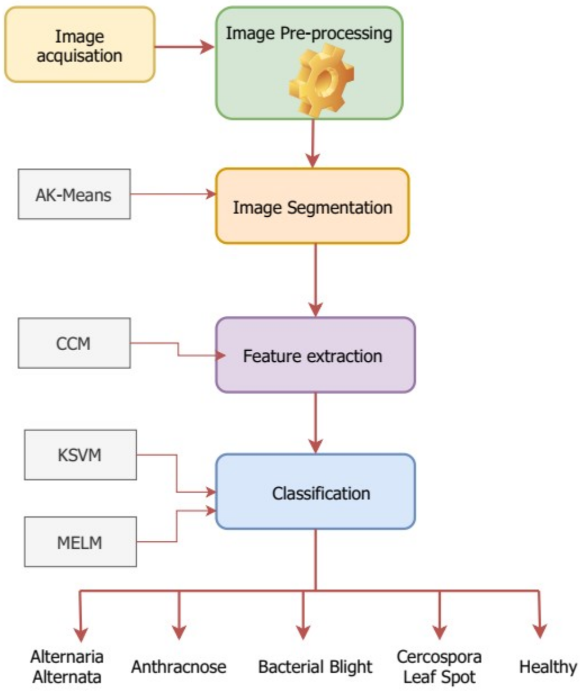


Fig. 4. Plant Leaf Disease Detection Process.

M. Leaf Color Extraction using H and B Components

Using an anisotropic diffusion approach, the input image is improved to preserve information of affected pixels before colour is separated from the base [22]. H and B components from the colour space of HIS and LAB are considered to discriminate between the grape leaf and the non-grape leaf.

N. Classification

After extracting features, the learning images from the learning database are classified using two separate machine learning algorithms: support vector machine and extreme learning machine. In addition, the SVM was used with two different kernels, linear and polynomial.

O. Support Vector Machine

The SVM supports a binary classifier that falls under machine learning. SVM is a binary classifier. The learning model is supervised and examines classification and regression data. Consequently, the most efficient hyperplanes selected for the SVM Classifier have been chosen, separating every input sample into two classes based on the two-class binary classification issue.

$$X = \left\{ \left(y^1, z^1\right), \left(y^2, z^2\right), ., \left(x^n, y^n\right) \right\} \tag{9}$$

The classification line is derived as:

$$X\left(n\right) = \left(w, n\right) + a \tag{10}$$

where z denotes X class label, and n is the high dimensional space-represented sample vector. The equation of the classification line is achieved with the help of w and a parameter. This work used 150 images of 5 different kinds of leaves to develop the SVM-based supervised machine learning classification with two different kernels. The test images of the NodeMCU and the camera were used to investigate the performance of the specified classifiers after the training procedure.

P. Proposed Multihidden-Layer ELML (MELM)

Fig. 5 illustrates the structure of the MELM (choose, for instance, the 3-level hidden layer ELM). Fig. 6 shows the workflow of the 3-level hidden layers using the MELM approach. To train the network, we provided the samples $\left\{X, T\right\} = \left\{xi, ti\right\}\left(i = 1, 2, 3, ..., Q\right)$ and constructed the network topology with hidden layers $\left(therearethreelayersineachofwhichhashiddenneurons\right)$ with activation function $g\left(x\right)$. Input layer, three hidden layers and output layer are available for the 3 level hidden layer structure of ELM. In this present work, we used three hidden layers placed with each other as two hidden layers. Thus the weight matrix β_{new} can be obtained from the second and output layers of the network. We can employ β weights, which enhance the overall ability of the network, depending on the number of actual samples. Then the MELM divides the previously combined three hidden levels and has three hidden layers for the MELM structure. This allows the predicted results of the third hidden layer $H3 = t\beta^+_{new}$ The β^+_{new} weight matrix is generalized inversely. The third MELM denotes the required matrix $W_HE1 = [B_2W_2]$ allows for the calculation of the formula (12) and formula $H3 = g\left(H_2W_2 + B_2\right) = g\left(W_{HE1}H_{E1}\right)$ acquired by the parameters of the third layer.

$$W_{HE1} = g^-1\left(H_3\right)H^1_{E1} \tag{11}$$

While H_2 denotes the second layer actual output, W_2 denotes the weight present in the second and the third hidden layer, B_2 denotes the third layer biasing factor, and H^+_{E1} denotes the inverse of the $H_{E1} = [1H_2]^T$ generation T, 1 is a one-column size Q vector with scalar unit elements 1. The $g^-1\left(x\right)$ notation denotes the opposite of $g\left(x\right)$ activation function. We use several activation functions for classification and regression to test the performance of the proposed MELM algorithm. We generally use the $g\left(x\right) = \frac{1}{(1+e^-x)}$ logistic sigmoid function. This determines the actual output of the third hidden layer:

$$H_4 = g\left(W_{HE1}H_{E1}\right) \tag{12}$$

Finally, the output new weights matrix β_new can be obtained from the third hidden and the output layer, the computation is as follows: If the number of hidden layer neurons is lower than the number of training samples β can be indicated as continues to follow:

$$\beta_{new} = \left(\frac{1}{\lambda} + H^T_4H_4\right)^- 1H^T_4T \tag{13}$$

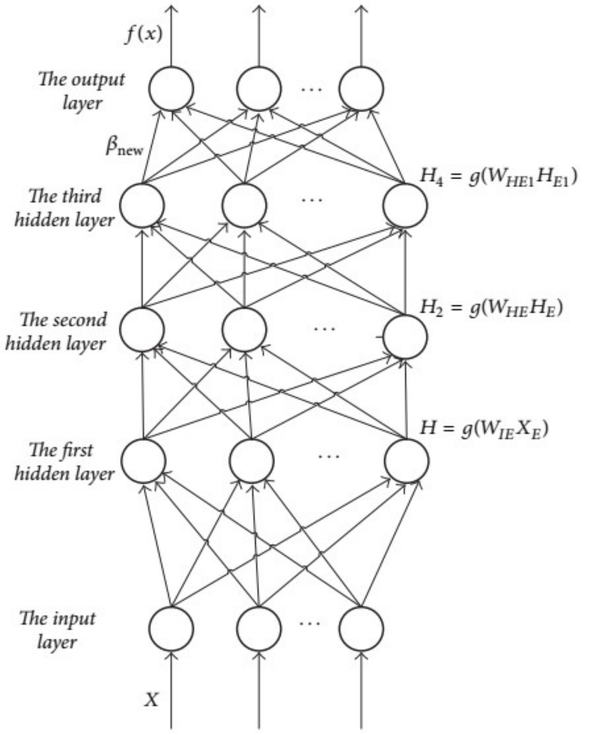


Fig. 5. Network Layers with 3-Level ELM

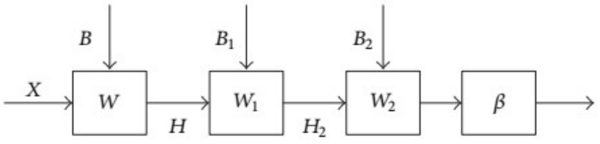


Fig. 6. The Workflow of the 3-Level Hidden-Layer of the ELM.

If the number of neurons in the hidden layer exceeds the number of training data, β is the following:

$$\beta_{new} = H_4^T \left(\frac{1}{\lambda} + H_4^T H_4\right)^{-} 1T \tag{14}$$

The actual output of the ELM network with three hidden levels can be described below:

$$f(x) = H_4 \beta_{new} \tag{15}$$

The operation phase is adjusting the network structural parameters from the second hidden layer to guarantee for Final hidden layer output be close to the estimated hidden layer performance for the total training time. Above is the estimated parameter for the ELM network of 3-level hidden layers, but our present work is intended to compute the ELM network variables from the multiple hidden layers and the final output of the MELM underlying network. The actual output of the ELM network of three hidden layers is as follows:

$$f(x) = H_4 \beta_{new} \tag{16}$$

The operation process optimises the network structure terms from the second hidden level to ensure that the actual hidden layer result is close to the estimated hidden layer output during the complete training time.

IV. Results and Discussion

In this study, the testing prototype and detecting plant leaf diseases are two phases of execution. The experimental prototype consists of the primary configuration and functioning of hardware components for necessary applications. Likewise, the analysis of plant leaf diseases detecting diseases from the images of the leaf dataset in question. There are two parts of system design, the primary is the sensor monitoring process, and the secondary is to predict the leaf diseases. The heart of the architecture is the NodeMCU microcontroller. The sensors are connected to the microcontroller, and gathered data is sent to the controller and received data will be further processed by the Matlab application for leaf disease detection. The sensors must capture the current environmental data. The early diagnosis of a disease depends on the image captured from the actual crop field using the MATLAB simulation tool. These images are the test cases for the identification process of plant disease.

A. Evaluation

The evaluation involves implementing a system of environmental monitoring sensors, including a software package required for deployed VGA cameras in the crop field to have precise agriculture for farmers. Database Description: The databases comprise 237 plant disease images from the leaf. The two most excellent plant disease image database websites collect five categories of images affected. Arkansas Plant Database and Reddit Plant Leaf Data Sets [14] [15] are used to acquiring images of diseases affected. Fig. 7 shows sample images of each form of plants leaf disease.

B. Evaluation of Segmentation

Two parts address the result section: first is the output of the segmentation of leaf diseases and the performance comparison of the segmentation in different parameters. The second part is detecting diseases from the collected images of the leaf dataset for the MELM and SVM classifications. For performance evaluation between the MELM segmentation output and the manual segmentation output, the Dice similarity coefficient (DSC), the mean square error (MSE) and the structural similarity index measure (SSIM) parameters were determined.

For the ten leaf samples, the average performance of the segmentation output, as shown in Table I, illustrates that the three parameters are acceptable for leaf disease detection.

TABLE I. Performance Parameter Values of the Segmentation Result

Disease Type	*DSC*	*MSE*	*SSIM*
Alternaria Alternata	0.97	0.014	0.97
Anthracnose	0.98	0.021	0.98
Bacterial Blight	0.95	0.047	0.96
Leaf Spot	0.96	0.011	0.94
Healthy	0.97	0.010	0.99

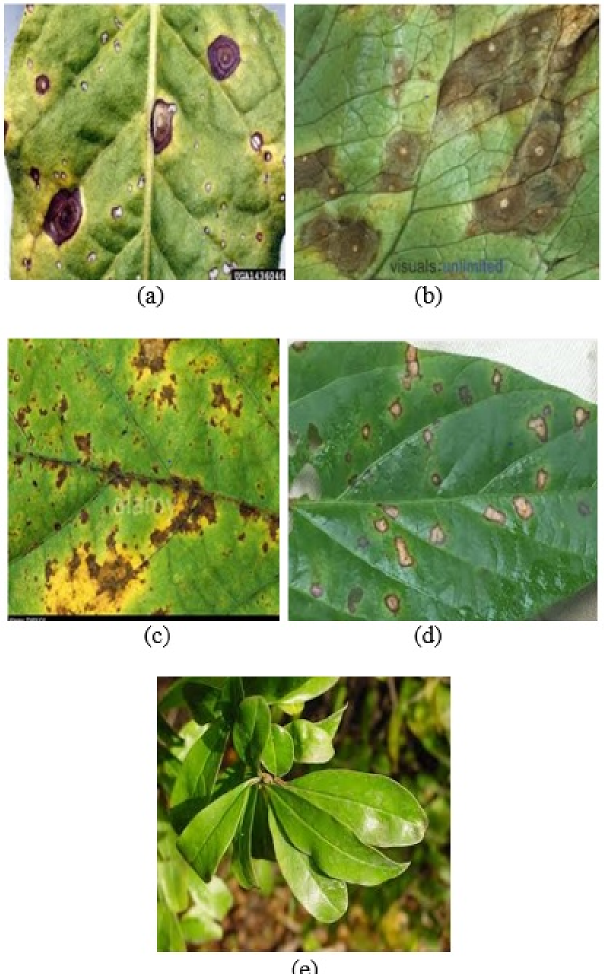


Fig. 7. Sample Images from the Database of Plant Disease: (a) Alternaria Alternata (b) Anthracnose (c) Bacterial Blight (d) Leaf Spot (e) Healthy

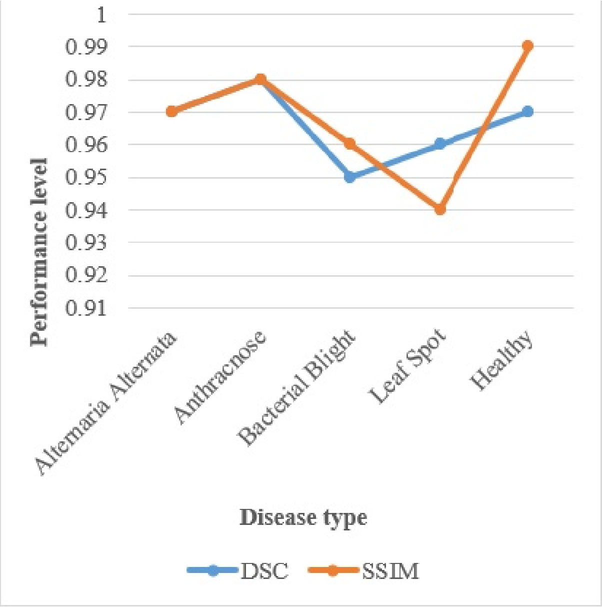


Fig. 8. Segmentation Result of the Parameters DSC and SSIM.

Fig. 8 shows that good segmentation with the proposed MELM techinique for five types of plant diseases having high DSC and SSIM values.

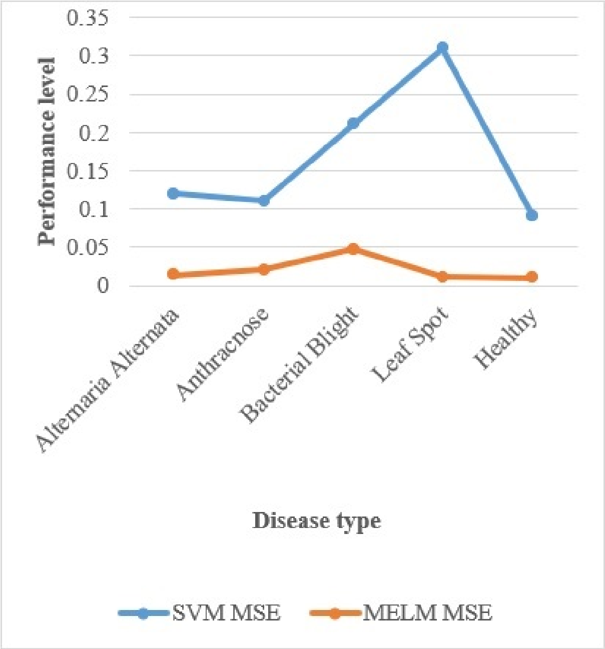


Fig. 9. Segmentation Result of the Parameter MSE of Two Classifiers

Fig. 9 illustrates good segmentation of five classes of plantation diseases with the minimum MSE values by the proposed MELM technique when SVM is poorly functioning owing to overfitting.Table II shows that the proposed MELM classification performs in classifying leaf diseases with other available approaches.

TABLE II. Performance Parameter Values of the Classification Result of Two Classifiers

Disease Type	*KSVM Accuracy*	*MELM Accuracy*
Alternaria Alternata	98.28	99.12
Anthracnose	97.32	98.92
Bacterial Blight	96.21	98.31
Leaf Spot	97.81	98.21
Healthy	98.21	99.14

The performance was evaluated of the two classifying models on the test data set, and the accuracy parameters for the performance assessment were calculated. The test data set includes 37 images with five different types of diseases. Fig. 10 demonstrates that the average five-class performance represents the average classification performance based on the test dataset after the individual classification performance is computed. The proposed multi hidden layers extreme machine classification delivers good performance of 99.12 in the classification of leaf diseases in comparison to the Support Vector Machine classification, which gives 98%.

V. Conclusion

The idea of detecting and monitoring leaf diseases of environmental conditions and irrigation systems is put into

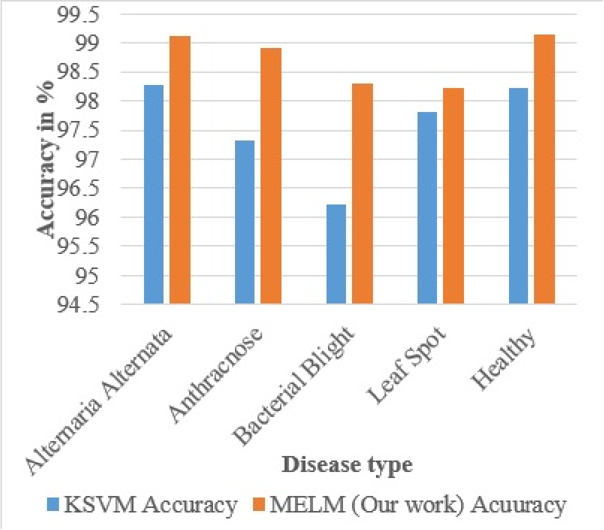


Fig. 10. Comparison of Classification Result of the Two Classifiers.

practice. For successful cultivation of crops, the correct detection and classification of the plant disease are crucial, and this may be done with image processing. This present study dealt explicitly with the adaptive K-Means clustering technique via various techniques for segments of the diseased part. This paper also explored several colour co-occurrence and classification methods for extracting the features of the diseased block and plant disease classification. The deployment of Multihidden Extreme Learning Machines can be efficiently employed for the classification of diseases in plants. The overall accuracy is considerably more significant for the multiple classifications than the ELM network structure. The MELM improves the network structure's performance in certain instances. In future, the data collected from Unmanned Ground Vehicles and Unmanned Aerial Vehicles can be applied to different ensemble algorithms.

References

[1] S. Monteleone, E. Moraes, B. Tondato de Faria, P. Aquino Junior, R. Maia, A. Neto, and Toscano, "Exploring the adoption of precision agriculture for irrigation in the context of agriculture 4.0: The key role of internet of things. sensors," *MDPI*, vol. 20, no. 10.3390/s20247091, pp. 46–129, 2020.

[2] S. Askraba, A. Paap, K. Alameh, J. Rowe, and C. Miller, "Laser-stabilized real-time plant discrimination sensor for precision agriculture," *MDPI*, vol. 161-1, no. 1-1. 10.1109/JSEN.2016.2582908, pp. 46–129, 2016.

[3] M. Rama, "Precision agriculture using iot," *International Journal for Research in Applied Science and Engineering Technology*, vol. 9, no. 122-127. 10.22214/ijraset.2021.36255, pp. 46–129, 2021.

[4] R. John, George G, N. Thomas, and R. Mammutil, "Application-specific wsn for precision agriculture," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 241-245. 10.1109/ISED.2018.8704076, pp. 46–129, 2018.

[5] Q. Jin, H. Liu, C. Wang, X. Wang, Q. Min, W. Wang, J. Sardans, X. Liu, X. Song, X. Huang, and J. Penuelas, "Greenhouse gas emissions in a subtropical jasmine plantation managed with straw combined with industrial and agricultural wastes," *Cambridge University Press*, vol. 56, no. 2, p. 280–292, 2020.

[6] Al-Hiary, Heba, Bani-Ahmad, Sulieman, Ryalat, Mohammad, Braik, Malik, Alrahamneh, and Zainab, "Fast and accurate detection and classification of plant diseases," *International Journal of Computer Applications*, vol. 517, no. 10.5120/2183-2754, pp. 46–129, 2020.

[7] S. Mohanty, D. Hughes, and M. Salathe, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, no. 10.3389/fpls.2016.01419, pp. 46–129, 2016.

[8] S. e. A. Raza, G. Prince, J. Clarkson, and N. Rajpoot, "Automatic detection of diseased tomato plants using thermal and stereo visible light images," *Frontiers in Plant Science*, vol. 10, no. e0123262. 10.1371/journal.pone.0123262, pp. 46–129, 2015.

[9] G. Owomugisha, J. Quinn, E. Mwebaze, and J. Lwasa, "Automated vision-based diagnosis of banana bacterial wilt disease and black sigatoka disease," *Frontiers in Plant Science*, vol. 10, no. e0123262. 10.1371/journal.pone.0123262, pp. 46–129, 2019.

[10] P. A, S. Chakraborty, A. Kumar, and O. Pooniwala, "Intelligent crop recommendation system using machine learning," *Frontiers in Plant Science*, vol. 10, no. 843-848. 10.1109/ICCMC51019.2021.9418375, pp. 46–129, 2021.

[11] J. Ma, K. Du, F. Zheng, L. Zhang, and Z. Sun, "A segmentation method for processing greenhouse vegetable foliar disease symptom images," *Information Processing in Agriculture*, vol. 6, no. 2, pp. 216–223, 2019.

[12] S. Khan and M. Narvekar, "Novel fusion of color balancing and superpixel based approach for detection of tomato plant diseases in natural complex environment," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, Part B, pp. 3506–3516, 2022.

[13] N. K. Trivedi, V. Gautam, A. Anand, H. M. Aljahdali, S. G. Villar, D. Anand, N. Goyal, and S. Kadry, "Early detection and classification of tomato leaf disease using high-performance deep neural network," *Sensors*, vol. 21, no. 23, pp. 1424–8220, 2021.

[14] R. Sagar Vetal, "Tomato plant disease detection using image processing," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 6, no. 6, pp. 2278–1021, 2017.

[15] S. K. Yusuke Kawasaki, Hiroyuki Uga and H. Iyatomi, "Basic study of automated diagnosis of viral plant diseases using convolutional neural networks," *Advances in Visual Computing*, pp. 638–645, 2015.

[16] C. M. B. Prema K, "Smart farming iot based plant leaf disease detection and prediction using deep neural network with image processing," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 9, pp. 2278–3075, 2019.

[17] P. J. Prof Madhavi Patil, Gaurav Langar and N. Panchal, "Tomato leaf disease detection using artificial intelligence and machine learning," *International Journal of Innovative Technology and Exploring Engineering*, vol. 5, no. 7, pp. 2456–0774, 2020.

[18] l. B. kota Sandeep, Gadde Divya and paduchuriRohithVenkatpavan, "Leaf disease detection system using raspberry-pi using neural network," *Journal of Emerging Technologies and Innovative Research*, vol. 7, no. 4, pp. 2349–5162, 2020.

[19] T. Roopali Gupta, "Tomato leaf disease detection using back propagation neural network," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 8, pp. 2278–3075, 2020.

[20] K. Thangadurai and K. Padmavathi, "Computer vision image enhancement for plant leaves disease detection," *Frontiers in Plant Science*, vol. 10, no. 173-175. 10.1109/WCCCT.2014.39, pp. 46–129, 2014.

[21] S. Phadikar and J. Sil, "Rice disease identification using pattern recognition techniques," *ICCITECHN*, vol. 10, no. 420 - 423. 10.1109/ICCITECHN.2008.4803079, pp. 46–129, 2009.

[22] A. Meunkaewjinda, P. Kumsawat, K. Attakitmongcol, and A. Srikaew, "Grape leaf disease detection from color imagery using hybrid intelligent system," *ICCITECHN*, vol. 10, no. 1. 513 - 516. 10.1109/ECTICON.2008.4600483, pp. 46–129, 2008.

Learning Global Average Attention Pooling (GAAP) on Resnet50 Backbone for Person Re-identification Problem

Syamala Kanchimani, Maloji Suman, P. V. V. Kishore
Department of Electronics and Communication Engineering,
Koneru Lakshmaiah Education Foundation,
Vaddeswaram, India

***Abstract*—Person re-identification has been an extremely challenging task in computer vision which has been seen as a success with deep learning approaches. Despite successful models, there are gaps in the form of unbalanced labels, poor resolution, uncertain bounding box annotations, occlusions, and unlabelled datasets. Previous methods applied deep learning approaches based on feature representation, metric learning, and ranking optimization. In this work, we propose Global Average Attention Pooling (GAAP) on Resnet50 applied on four benchmark Re-ID datasets for classification tasks. We also perform an extensive evaluation on the proposed Attention module with different deep learning pipelines as backbone architecture. The four benchmark person Re-ID datasets used is Market-1501, RAiD, Partial-iLIDS, and RPIfield. We computed cumulative matching characteristics (CMC) and mean Average Precision (mAP) as the performance evaluation parameters of the proposed against the state of the art. The results obtained have shown that the added attention layer has improved the overall recognition precision over the baselines.**

***Keywords*—Person re-identification; attention network; ResNet50; global average attention**

I. Introduction

The goal of person re-identification (Pe-reID) is to identify and fetch a random person across mutually exclusive camera sources [1]. The objective of Pe-reID models is to determine whether a given query person has reappeared in the frame at a different point in time or in any other camera source at the same point in time [2]. The given query of the person can be an image [3], video [4] and also in text format describing some attribute of the query person [5]. The application range of Pe-reID spans surveillance systems with intelligence that can provide automated feedback on people's movements in real-time.

The Pe-reID pipelines are made up of 5 different tasks. They are arranged chronologically as data collection, bounding box creation, training data annotations, model design and person re-identification. According to [6], the above steps are being considered as a closed world Pe-reID system where the data is structured effectively during preparation. Whereas the open-world Pe-reID system will operate on raw datasets with no annotations and labels. Specifically, this work uses the closed world approach to the Pe-reID problem. The closed world setting is based on the following conditions:

1) Single modality video or image data has been used.
2) The annotations are fixed with persons in bounding boxes with same area identities.
3) The query person is extracted from the training data.
4) Finally, there is enough training data from annotations for supervised learning of person re-identification.

The above processes require expertise and domain knowledge for transforming a video surveillance security problem into a challenging person re-identification problem.

Pe-reID is still considered a super constrained problem. This is due to multiple challenges such as background clutter, low image resolution, poor image quality, partial occlusions, uneven bounding box annotations, human-object interactions [7], etc. Preliminary investigations on Pe-reID problems focused on the hand-crafted feature extraction methods such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). Similar Works considered body reconstruction models and distance metric learning [8]. In the last decade, the growth of AI-based approaches has captured the Pe-reID problem and has subsequently proved its potential with exceptionally good recognition accuracies across datasets [9]. However, the results obtained were nowhere close to the requirements of a real-time deployment pipeline.

In this work, we propose to redesign the regular deep learning approaches with additional layers to learn focused attention. The proposed attention model is called the Global Average Attention Pooling (GAAP) network. The GAAP learns by averaging the features extracted from previous convolutional layers in query, key and value channels. The output of the GAAP network is used to select features that contribute to making correct decisions about a given query person. The attention block has been created based on the non-local attention technique from [2] and the global average pooling is initiated on the attention features to generate a maximally discriminating learnable feature representation.

The proposed GAAP layers or block is integrated with the existing benchmark deep networks such as VGG, ResNet and Inception Net. All these models are trained from scratch on four different types of Pe-reID image datasets. The evaluation metrics used are Cumulative Matching Characteristics (CMC) and mean Average Precision (mAP). The proposed GAAP block is integrated at the last layer of the backbone networks as against the previous works where it was added in the primary layers. Experiments are designed to test and evaluate the GAAP block's performance in the identification of people in the test data.

The proposed GAAP integrated framework differs in three major areas from the existing baselines:

1) The addition of an attention-based GAAP layer after the convolutional layers in the backbone network ensures a maximally discriminating of learnable features for the dense layers.
2) The global average pooling is performed to mobilize feature space obtained from a large set of convolutional filters into a singular feature representation.
3) An attempt is made to validate the proposed GAAP integrated deep learning framework with cross dataset testing.

To develop the proposed GAAP block for Pe-reID problem, we propose the following objectives:

1) To restructure the datasets into training, testing and validation sets as input to the backbone networks with an integrated GAAP block.
2) To train the deep learning pipelines with an integrated GAAP layer and evaluate their performance.
3) To construct experiments for validating the proposed model.

The rest of the paper is organized as follows. The next section presents previous research performed on person re-identification problems. The third section gives an elaborated discussion on the construction, training and testing of the integrated GAAP network with backbone architectures. Experiments were built for validation and the obtained results are analyzed expensively for characterizing the model pipelines for Pe-reID. Finally, Section 5 concludes on the attained research outcomes with future direction.

II. Literature Review

Remarkably, the closed world Pe-reID problem [6] has been acknowledged previously through the following compositions: 1) Feature representational learning, 2) metric learning and 3) ranking optimization. In this section, we discuss the above approaches and their underlying research findings with their capabilities for providing better Pe-reID solutions. Feature representation learning has been implemented with three derivatives such as global, local and auxiliary. The global features represent the whole person's image for learning [10]. Contrastingly, the local feature learning approaches use parts of the image as input to the extraction algorithm [11]. However, in auxiliary feature representation, data generation models such as Generative Adversarial Networks (GAN) are used to learn variations in the existing datasets. The features of the entire person image are learned in the global feature representation model. These models used powerful deep learning architectures as a classifier for the identification of persons. A set of highly discriminating features were captured with single image representation and cross image representation on triplet loss embedding [12]. The other most popular Pe-reID models treated it as a multi-class classification problem [13] and multi-scale representation problem [14]. Though global feature representation learning has leveraged its full potential for giving good accuracies, it suffered from overfitting problems.

The overfitting problem occurred in global feature representation as the network learned mostly the background information rather than focusing on the person of interest. The global features also have an image misalignment problem that is induced because of the multiple views and person orientations in the training images. Part-based local feature representation has been proposed to overcome the misalignment and overfitting problems in global features. Two mechanisms were formulated in the form of pose-based [15] and rough horizontal-based [16] body part detection for training. In the automated body part detection models, the full body and part features were fused together for classification. Especially, part base local feature representation models such as multi-channel aggression [13], multi-scale context-aware convolutions [14], multi-stage feature decomposition [2], and bilinear pooling [3] have shown expedient performance. Moreover, the performance has been enhanced further by pose-driven [17] and pose-guided matching [18] methods. However, in horizontally divided part-based models, the part-based conventional baseline [18] has served as a building platform for part-based Siamese long short-term memory (LSTM) networks [3]. Other highly accurate models such as Interaction and Aggregated (IA) [19], and second-order nonlocal attention [3] have used reinforced feature learning approaches. The local feature representation learning approaches are limited by the use of noisy pose estimations and large background clutter. Some of the problems associated with both global and part-based feature learning models were maneuvered efficiently by using additional attributes in the training data. These additional attributes are generative datasets [3], semantic representations [20], viewpoint data [17] and data augmentation [18]. The above auxiliary features are found to provide additional data samples for training, which greatly enhanced their ability to identify persons. However, these auxiliaries are computationally expensive and required an additional pre-processing stage for input pairing.

Apart from image-based Pe-reID methods, some recent works have used video-based inputs to relocate a person in the multi-view video frames. Though the video representation has more information in the form of both spatial and temporal data, they fail to capture them accurately due to the unpredictable nature of the persons appearing in the video sequences. Predominantly, recurrent neural networks (RNN) were proposed to capture the temporal information [21] with a temporal pooling layer at the end of RNN. Mixed attributes of spatial and temporal information using sequential fusion are used to enumerate the frame-level feature representations for improved recognition [22]. A varying length video sequence is considered challenging in most video-based applications. In [19], long video sequences are divided into tiny snippets and are ranked in descending order to learn the compact embedding from the top – K segments.

In most of the works on Pe-reID, the backbone architecture is similar to that of the standard ones used for image classification tasks such as VGG-16 and ResNet50. Few works on Pe-reID have modified the ResNet backbone by introducing size 1 in the last convolutional layer or by adding adaptive average pooling in the last pooling layer [23]. However, a tremendous amount of design time can be curtailed by adopting AutoML models for Pe-reID as shown in [20]. The primary objective of all the above-discussed models is to improve the identification accuracy of the person. One such model which has improved the performance of the Pe-reID deep learning methods is deep metric learning (DML) [24]. The DML uses a metric loss function to calculate the distance between the features from within the class and between classes during training for generating a maximally discriminant feature vector for classification in the dense layers. The identity loss has been widely studied in multiple Pe-reId methods than any other models. The other type of DML model that has indeed improved the performance of the Pe-reID is triplet loss embedding, which starts by computing the distance between the positive class pairs and negative class pairs. The learning is initiated by maximizing the distance between the negative pairs and minimizing it between the positive pairs. The only shortcoming is during the pre-processing stage where the pairing process is performed between the samples from within the class and across classes. Moreover, this pairing complexity increases with the increase in the number of samples per class or an increase in the number of classes itself. Similar to the above DML model, ranking optimization has been shown to improve retrieval efficiency during the testing phase [25]. Very recently, attention-based models [2] have been shown to further strengthen the efficiency of Pe-reID models.

In this work, we propose a global average attention pooling (GAAP) layer at the end of the convolutional layers in ResNet backbone architectures and evaluate its performance against state-of-the-art models. We evaluate the importance of the proposed GAAP against various attention models and across two popular backbone architectures VGG and ResNet. Finally, we conclude by reasoning the significance of the GAAP layer in Pe-reID implementation through experimentation.

III. Methodology: Rank View Triplet Loss Embedding

Learning in Pe-reID is accomplished with D data samples $X_{Pe-reID} = \{x_i, y_i\}\ \forall\ i = 1\ to\ D$ with the goal of finding a mapping function between input x_i and their corresponding labels y_i. The objective of the Pe-reID deep learning neural networks is to learn a mapping function $\theta : X_{Pe-reID} \mapsto F$ that transforms the combined feature space $X_{Pe-reID}$ into F, in which the samples are highly discriminative. Given a set of test images T_{Pe_reID}, the learned mapping function θ will try to project the test images into constituent labels. The testing images and training images in this case are totally nonoverlapping. The primary challenge in the above model is in the learning process of the mapping function θ which in the previous works has been a simple image classifier. The mapping function in case of Pe-reID has to adapt to varying and insufficient training samples per class which results in inconsistent loss parameters during the training process. To regularize the loss function during training we propose to induce an attention layer with global averaging pooling as an architectural upgrade to the existing 50-layer ResNet model. In this subsection, we present the complete Global Average Attention Pooling (GAAP) network and deconstruct the entire pipeline implementation in tensorflow2.3.

A. GAAP Architecture

Global Average Attention Pooling (GAAP) is a network built on top of the existing ResNet50 model with an additional 4 layers. The first three layers in GAAP are convolutional layers, and the last one is a pooling layer. The overall GAAP model for Pe-reID has been illustrated in Fig. 1. A multiscale architecture such as ResNet50 is being used as the backbone network as it has become the state-of-the-art model in many previous works [14]. The proposed attention model has effectively shown to fuse the low level and high-level features to isolate the features of importance within a class label. The attention features are further averaged globally to regenerate a highly discriminative feature map for a particular class label. The model is end–to–end trained on the classification loss function only, which is categorical cross-entropy.

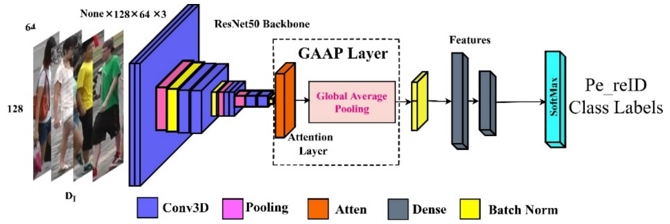


Fig. 1. Global Average Attention Pooling Network Illustration for Person Re-Identification.

The mark of a good person reidentification model is to retrieve accurately the query person image that closely matches the sample images in the class label. However, the previous models used different loss functions to attain close matches between the quey and training images. The two most commonly used loss functions are contrastive and triplet loss. The implementation of these loss functions requires enormous computation resources for training. The proposed GAAP attention framework adds only four layers to the existing network and therefore occupies uses comparatively lesser computational resources. Moreover, the total parameters of the GAAP architecture are lesser than the metric learning models.

The model in Fig. 1 takes input images in training data and transform them into features. The backbone network is Resnet with 50 layers with skip connections.The appearance images $A_n(x)\ \forall\ n = 1\ to\ N$ is divided into a length of N samples per class, the appearance sequence is a multidimensional tensor represented as $\mathrm{A} \in R^{r \times c \times 3}$. Here, (r, c) are RGB image height and width in three color channels. Since the GPU capacity is 8GB, the images are standardized to 128×64×3 across all datasets. This becomes the input to the RGB appearance stream $S_A \rightarrow A_n(None, x, y, c)$. The S_A stream is made from backbone ResNet consisting of multiple convolutional, maximum pooling with rectified linear activations and batch normalization layers. There is no padding in convolutional layers. This S_A stream will extract features from A using the trainable parameters Θ_{S_A} by optimizing the loss function L_{S_A} on the entire dataset

$$\Theta_{S_A} = \arg \min_{\Theta_{S_A}} L_{S_A}\left(\Theta_{S_A} : \mathrm{A}\left(x\right), y\right) \tag{1}$$

Here y denotes the class labels. The S_A stream is optimized using the categorical cross-entropy loss L_{S_A} defined as

$$L_{S_A} = -\sum_{i=1}^{C} \left(y_i \times \log\left(y_i\right) + \left(1 - y_i\right) \times \log\left(1 - y_i\right)\right) \tag{2}$$

The trained model M(Θ_{S_A}) will output at the end of i^{th} convolutional layer with j^{th} appearance feature map by using the expression

$$F_A^{ij}\left(x,y\right)=f_a\left(\sum_{p}\sum_{n=0}^{r-1}\sum_{m=0}^{c-1}\left(W_{ijp}^{nm}*\mathrm{A}_{(i-1)p}\left(x+n,y+m\right)\right)+b_{ij}\right)\quad(3)$$

Where, A is the person image and f_a is the activation function. W_{ijp}^{nm} are the weights at position (n,m) associated with p^{th} feature map in the $(i-1)^{th}$ layer of the CNN ResNet50 network. The parameter b_{ij} is the bias associated with each of the neurons. Eq'n (3) depicts the convolutional operation between the images and the weight matrix, which is updated sequentially during training of the network. The output RGB appearance features has the dimension $F_\mathrm{A}\in R^{r_j\times c_j\times 3\times C}$. Here, C is the channels or filter kernels applied in j^{th} convolutional layer. These features are further processed using the attention layers before being applied to the dense layers for classification.

B. Attention Layers and Global Average Pooling (GAAP Attention Module)

The proposed attention layers are shown in Fig. 2. The proposed model is inspired by self-attention in [11]. It consists of four 1×1 convolutional layers with stride 1 and one residual connection to preserve the original feature encodings. The dot product enhances the features that are important and discards the others that are least useful in the decision process. This allows the features to concentrate on the areas of the pixels that are highly discriminative in nature. The difference between self-attention in [12] and the proposed in Fig. 2 is that the latter takes input from different features within the class for computing the attention maps. Contrastingly, the self-attention model uses the same features of a single sample to calculate the attention map. The attention map in out proposed model is calculated between the $F_A^i\left(x,y\right)$of i^{th} feature of an image A_i in a class and the $F_A^j\left(x,y\right)$ of the j^{th} feature of the same image in the class. These features are obtained from the learned backbone network. This enables the network to learn similar appearances across the same image with different features computed using the learned filters in the feature mapping network. The proposed cross-feature attention (CFA) is defined as

$$CFA\left(f_i\right)=soft\max\left(\frac{\alpha Q_i.K_i^T+\left(1-\alpha\right)Q_jK_j^T}{\sqrt{d_k}}\right).V_i\quad(4)$$

Where $i,j\in 1,...,J$, with J is the number of filters in the convolutional layers and $\alpha=0.5$ is the set hyperparameter for all the layers. The (Q,K,V) are the query, key and value as three convolutional layers in the Fig. 2.

The dot product enhances the features that are important and discards the others that are The weighted sum of features are obtained from all possible positions using the following learning model.

$$a_i=W_a\times\theta\left(f_A^i\right)+f_A^i\quad(5)$$

Where, a_i is the attention maps obtained from the learned W_a with parameters θ of the network. The $+f_A^i$ is the residual connection. Finally, to capture the domain specific features, we apply global average pooling instead of maximum pooling used regularly. The global average pooling of attention features is formulated as

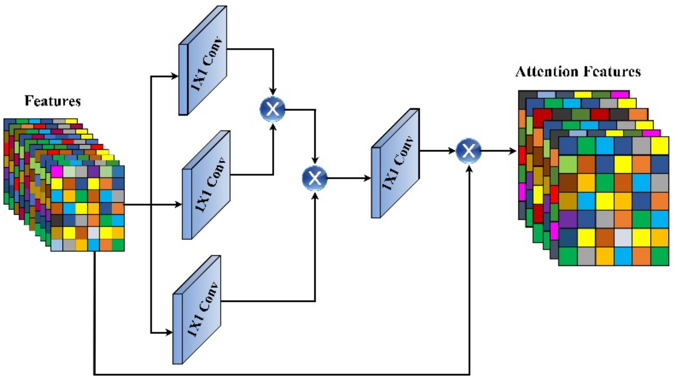


Fig. 2. Attention Layers and its Architecture.

$$f_{ga}=\left[f_1,f_2,......,f_K\right]^T=\left(\frac{1}{|F_k|}\sum_{f_i\in F_k}f_i\right)\quad(6)$$

Where, F_K is the total number of features in the feature maps with K features. f_k represents feature maps which are learned by the backbone and attention layers in the PereID pipeline using the backpropagation algorithm. Finally, the backbone Resnet50 is presented in Fig. 3.

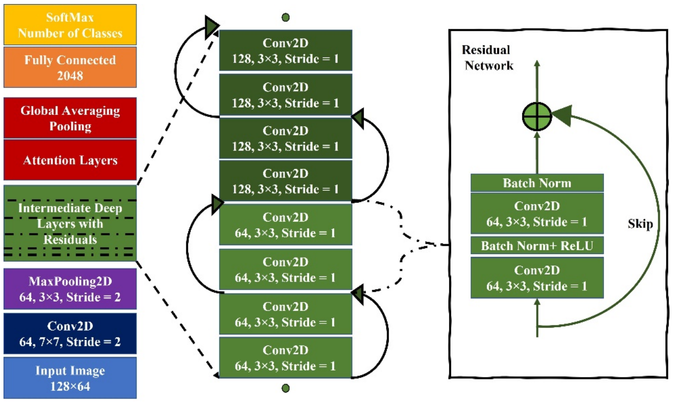


Fig. 3. The ResNet50 Architecture used as Backbone Network in the Proposed GAAP Model

Finally, the combined loss of the entire GAAP network is computed as

$$L_{GAAP}=\sum_{i=1}^{itr}L_{S_A}\left(\Theta\left(A_i\left(x\right);\theta\right),y_i\right)\quad(7)$$

Where, Θ is the trainable parameters of Resnet50 and θ are the learnable parameters of the attention network. The final feature representation is learned by minimizing the categorical cross entropy function L_{S_A} for classification. During testing only $A_i\left(x\right)$ are used for inferencing the trained model.

IV. Results and Discussion

The four benchmark person Re-ID datasets used in this work are Market-1501, RAiD, Partial-iLIDS and RPIfield. We evaluated the performance of the proposed method using the following parameters, computed cumulative matching characteristics (CMC) and mean Average Precision (mAP). This section gives details about the benchmark person re-identification datasets used for evaluation, the model configuration set for training and testing with in-depth assessment of attention framework.

A. Datasets for Pe-reID

This work has conducted extensive experimentation on four popular benchmark datasets, Market-1501 [1], RAiD [26], Partial-iLIDS [16] and RPIfield [27]. Market-1501 is the largest person re-identification image dataset containing 1501 identities captured with six camera angles and 32,668 bounding box persons that are annotated with deformable part model pedestrian detector. An average of 3.6 images are obtained per person per viewpoint. The training and testing sets have 750 and 751 classes respectively with 3368 additional query images. In this work, we train the model with 750 image classes and test with 750 classes. The training set is split into 15% validation. The image resolution is 128×64. Few training samples of the Market – 1501 are shown in Fig. 4. RAiD is developed in 2014 which has multiple person trajectories recorded using four static camera views. The data is primarily focused on persons on sidewalks and crosswalks. The images in the dataset appear cleaner in the background when compared to the other datasets used in this work. The RAiD dataset has 43 classes with 6290 image samples that are split into 0.7:0.15:0.15 for training, validation and testing. The image resolution is 128×64. Partial iLIDS has occluded person re-identification samples from 476 images from 119 classes. It contains four camera views with a varying resolution of the hand cropped images from the surveillance video data. However, the proposed work has set the resolution of all the images in the dataset as 128×64×3. The occlusions in the images are due to another person or luggage. The RPIfield is constituted in 2018 with 112 class identities with 12 non-overlapping camera viewpoints having 601581 samples is being shown in Fig. 4. The images are annotated using fast pyramid features for bounding box detection which is the reason for multiple resolutions across the dataset. However, the proposed work has normalized the use of image resolution to 128×64 across all the datasets and subsequently across the models used in this work.

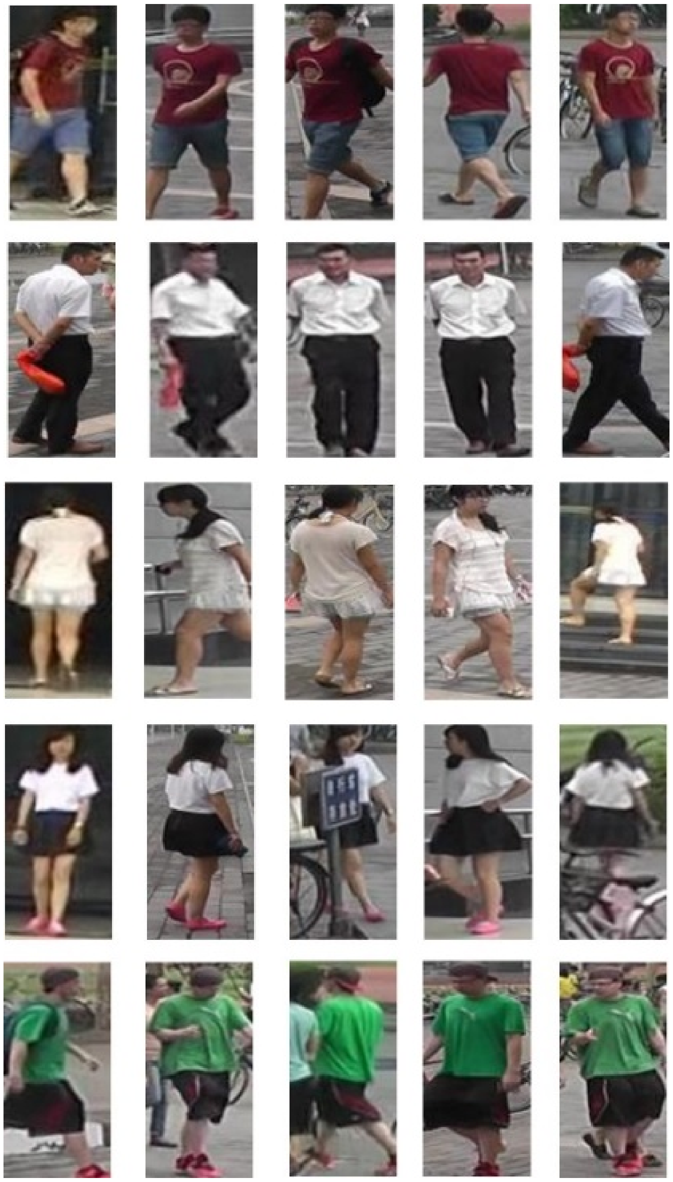


Fig. 4. Samples from Market – 1501 Dataset.

B. Model Configuration

For feature extraction we selected three Resent models as the backbone for feature extraction. The first was tiny ResNet-18 with the attention layers added before the dense layers. This model was used to train the model in a lightweight configuration for real time implementation. The feature extraction process was handled with the help of 8 convolution layers in ResNet-18. Similarly, we applied ResNet-34 with 32 and ResNet-50 with 48 convolutional layers each for feature extraction respectively. The ResNet-50 is deep with added 1×1 convolutions to preserve the input features and decrease the dimensionality of the feature vector. We also included the popular VGG – 16 and – 19 models to analyse the real time deployability as these models are highly recommended in this regard. To evaluate our proposed ResNet with attention layers, we adopted categorical cross entropy loss for optimization with Adam optimizer. This has resulted in providing level playing comparisons with the previous works. In the next subsection, evaluation protocols for the proposed model are being formulated.

C. Model Evaluation Protocols

All the backbone networks and the associated layers are trained from scratch on the benchmark datasets used in this work. Weights and biases were initialized using the standard zero mean 0.01 variance gaussian distribution function at the start of every training session. The network learns by updating weights and biases by optimizing the gradient losses that are backpropagated in reverse. The other training initialized hyperparameters would be learning rate, activations, momentum factor, frame dimensions, number of epochs, learning rate decay and minimum allowable loss of the trained classifier. The training set in each class is unbalanced in all the datasets and no attempt has been made to normalize the sample images in each class. However, data augmentation has been initiated on each image to increase the size of the dataset. Four types of augmentation were applied in the form of horizontal and vertical shifts, zoom and crop as shown in Fig. 5. The batch size was selected as 32 which means that there will be 32 images per training batch in each episode. The GAAP model with ResNet – 50 backbone was initialized with a learning rate of 0.0001 with a decay of 10% whenever the validation loss became constant for more than 4 epochs. The other backbones of ResNet were initialized with a higher learning rate to compensate for the lesser depth in features. The momentum factor has been considered as 0.8 across all networks and databases. The average number of training epochs were set at 25. Since the structure of the datasets has been unbalanced in the sample size, the image resolution is kept constant at 128×64 along with all other training initializers to maintain balanced evaluation.

Unseen training samples were used for testing the proposed GAAP network. The SoftMax outputs provide a statistical measure of the probability distribution of the test person image that closely matches the labels in the trained class. Here, we evaluate the global average pooling network with attached attention layers and their impact on the overall performance of the network across all datasets. Finally, we compare the proposed GAAP model with other Pe-reID methods and also perform a detailed ablation study to analyse the behaviour of the model under various test loads. All the models were trained and test on 8GB Nvidia A-4000 series with 16GB memory. The implementation has been done in TensorFlow and Keras packages.

D. Evaluation of the Proposed GAAP Model

The evaluation of the proposed method is conducted by calculation of cumulative matching characteristics (CMC) and mean Average Precision (mAP) across the training dataset. We computed single- and five-fold cross testing on all the datasets. We designed five backbone architectures to evaluate the models performance in identification of a person under various circumstances. Table I provides the results on all benchmark datasets with five backbone networks: VGG-16, VGG-19, ResNet-18, ResNet-34 and ResNet-50. The larger versions of ResNet such as ResNet-101 and ResNet-150 were not trained due to the GPU hardware insufficiency.

All the backbone networks were trained on exactly similar protocols as discussed in previous sections. The above table shows that VGG has failed to take advantage of the attention layers attached after the feature extraction convolutional layers. The reason for underperformance by VGG when compared to ResNet is the missing residual connections in the former model. The residual connections make the ResNet models to avoid overfitting and vanishing or exploding gradients problems. As the networks get deeper, the deep layers may sometimes get zero gradients as input which contributes to faulty decision making on the class labels. The success of ResNet – 50 is attributed to the fact that network is made deeper by adding 1×1 convolutions that help increase the feature quality and reduce the dimensionality.

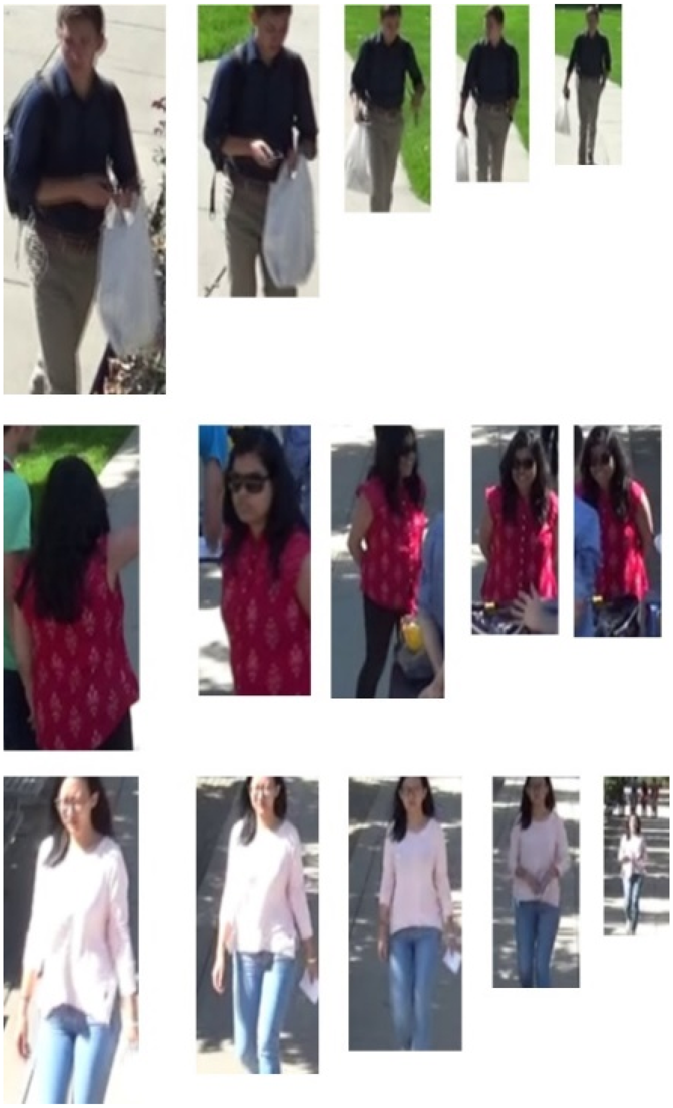


Fig. 5. Samples from RPIfield Pe-reID Dataset.

The results are as expected, and the overall recognition rate mRA with ResNet – 50 was averaged around 91.2% after 5-fold repetition. This is better than the other Pe-reID recognition frameworks in table I by an average of around 10%. The reason lies in the residual connections in ResNet - 50 and added attention module that highlighted the relationships between within-class samples to drive the appearance of the person in multiple cameras. RPIfield dataset has been shown to have maximum test accuracy due to the presence of large training data in all the considered datasets. In the next section, we evaluate the importance of the attention module.

E. Evaluation of Attention Layers

Table II shows the computed parameters on the test data with attention layers and without attention layers. The results show that the there is a 30% increase in network confidence for recognition with the proposed GAAP architecture when

TABLE I. Evaluation of the GAAP Model on Benchmark Datasets with Different Backbone Networks

Backbone Networks Trained in GAAP	Pe-reID Datasets	mRA		CMC	
		1 - Fold	5 - Fold	1 - Fold	5 - Fold
VGG - 16	Market-1501	0.731	0.776	0.727	0.746
	RAiD	0.743	0.788	0.732	0.752
	Partial-iLIDS	0.705	0.696	0.713	0.701
	RPIfield	0.741	0.81	0.764	0.775
VGG - 19	Market-1501	0.716	0.721	0.741	0.713
	RAiD	0.803	0.848	0.799	0.818
	Partial-iLIDS	0.815	0.86	0.804	0.824
	RPIfield	0.777	0.768	0.785	0.773
ResNet - 18	Market-1501	0.813	0.882	0.836	0.847
	RAiD	0.788	0.793	0.813	0.785
	Partial-iLIDS	0.84	0.885	0.836	0.855
	RPIfield	0.852	0.897	0.841	0.861
ResNet - 34	Market-1501	0.814	0.805	0.822	0.81
	RAiD	0.85	0.919	0.873	0.884
	Partial-iLIDS	0.825	0.83	0.85	0.822
	RPIfield	0.83	0.875	0.819	0.839
ResNet - 50	Market-1501	0.842	0.873	0.8	0.828
	RAiD	0.828	0.897	0.851	0.862
	Partial-iLIDS	0.803	0.808	0.828	0.8
	RPIfield	0.867	0.912	0.863	0.882

compared to the traditional models. All the models are trained using the same initial conditions as discussed in the Section 4.3. The big jump in the proposed model is due to the ability of the network to train on the features that are important for classification. The use of attention layers guarantees highly discriminative features within a class label. In the next section, we evaluate the global average pooing of attention features against the traditional maximum pooling model used in previous works.

F. Evaluation of Global Average Pooling

Before evaluating the global average pooling layers in GAAP architecture with backbone CNN models, we present the attention maps obtained on Market – 1501 dataset with ResNet – 50 backbone CNN model in Fig. 6. The figures provide a visual confirmation of the concentration of features used for training the dense layers in the GAAP pipeline for recognition of persons in Market – 1501 dataset. We observed similar kind of results across all other datasets used in this work.

In the following Table III, we computed the performance parameters CMC and mRA for the proposed global average pooling and maximum pooling of attention features after the convolutional layers in all the backbone networks used in this work. The results show that the global average pooling results in an 10 ±2% increase in performance of the backbone network for recognition of persons when compared to the traditional maximum pooling model. In the traditional maximum pooling model, the largest values in the feature space on the batch size within a class label are pooled together for training on the dense layers. This procedure generates outlier features that are not concentrated on the person or object of interest in the entire feature space at all times. Hence, it is intuitive that the final feature space for dense layers may possibly miss some of the prominent features necessary for correct identification. This can be avoided by considering an averaging feature space across a batch size within a class label. Consequently, the global averaged features have shown to exhibit the characteristics of all prominent regions of interest across a batch size given a generalized representation of the person across the class label. This is observable in the Fig. 7 visualization of attention regions projected on the original images of Market – 1501 dataset. Finally, we compare our proposed GAAP with ResNet – 50 backbone with the state – of – the – art methods for Pe-reID on the benchmark datasets in the following section.

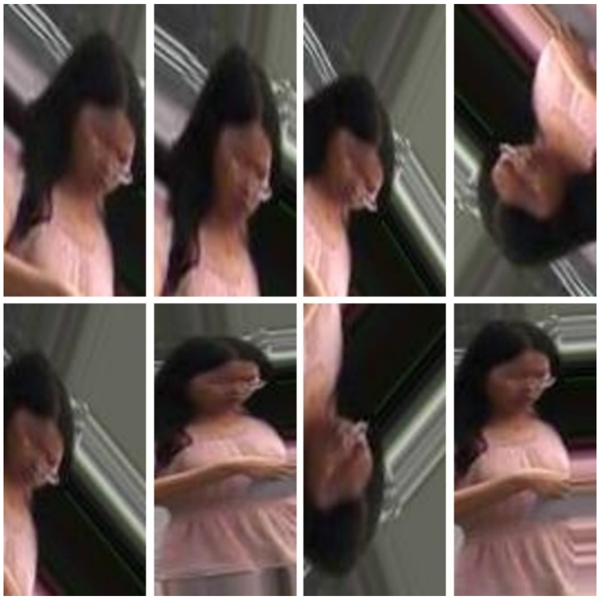


Fig. 6. Data Augmentation Applied on Market – 1501 Dataset.

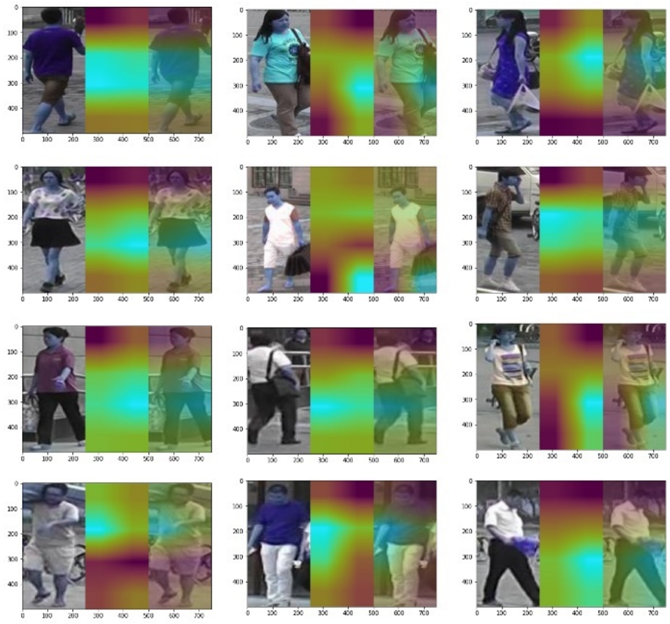


Fig. 7. Attention Maps Obtained from the Proposed GAAP Architecture on ResNet – 50 Backbone CNN Model.

G. Comparison with the State-of-the-Art Pe-reID Methods

This section draws comparisons of different Pe-reID methods against the proposed GAAP architecture. As can be observed from the above analysis that the ResNet – 50 backbone

TABLE II. Performance Evaluation of the Selected Backbone Networks on Pe-reID Datasets with and without Attention Layers

Classifiers	Datasets	With Attention Layers (GAAP)				Without Attention Layers			
		mRA		CMC		mRA		CMC	
		1 - Fold	5 - Fold	1 - Fold	5 - Fold	1 - Fold	5 - Fold	1 - Fold	5 - Fold
VGG - 16	Market-1501	0.731	0.776	0.727	0.746	0.504	0.489	0.497	0.499
	RAiD	0.743	0.788	0.732	0.752	0.546	0.544	0.569	0.539
	Partial-iLIDS	0.705	0.696	0.713	0.701	0.476	0.474	0.445	0.465
	RPIfield	0.741	0.81	0.764	0.775	0.54	0.546	0.574	0.564
VGG - 19	Market-1501	0.716	0.721	0.741	0.713	0.521	0.508	0.51	0.504
	RAiD	0.803	0.848	0.799	0.818	0.576	0.561	0.569	0.571
	Partial-iLIDS	0.815	0.86	0.804	0.824	0.618	0.616	0.641	0.611
	RPIfield	0.777	0.768	0.785	0.773	0.548	0.546	0.517	0.537
ResNet - 18	Market-1501	0.813	0.882	0.836	0.847	0.612	0.618	0.646	0.636
	RAiD	0.788	0.793	0.813	0.785	0.593	0.58	0.582	0.576
	Partial-iLIDS	0.84	0.885	0.836	0.855	0.613	0.598	0.606	0.608
	RPIfield	0.852	0.897	0.841	0.861	0.655	0.653	0.678	0.648
ResNet - 34	Market-1501	0.814	0.805	0.822	0.81	0.585	0.583	0.554	0.574
	RAiD	0.85	0.919	0.873	0.884	0.649	0.655	0.683	0.673
	Partial-iLIDS	0.825	0.83	0.85	0.822	0.63	0.617	0.619	0.613
	RPIfield	0.83	0.875	0.819	0.839	0.633	0.631	0.656	0.626
ResNet - 50	Market-1501	0.842	0.873	0.8	0.828	0.563	0.561	0.532	0.552
	RAiD	0.828	0.897	0.851	0.862	0.627	0.633	0.661	0.651
	Partial-iLIDS	0.803	0.808	0.828	0.8	0.608	0.595	0.597	0.591
	RPIfield	0.867	0.912	0.863	0.882	0.64	0.625	0.633	0.635

TABLE III. Comparative Analysis of Global Average Pooling and the Traditional Maximum Pooling of Attention Features for Pe-reID Tasks

Classifiers	Datasets	With Global Average of Attention Features (GAAP)				With Maximum Pooling of Attention Features			
		mRA		CMC		mRA		CMC	
		1 - Fold	5 - Fold	1 - Fold	5 - Fold	1 - Fold	5 - Fold	1 - Fold	5 - Fold
VGG - 16	Market-1501	0.731	0.776	0.727	0.746	0.628	0.613	0.621	0.623
	RAiD	0.743	0.788	0.732	0.752	0.67	0.668	0.693	0.663
	Partial-iLIDS	0.705	0.696	0.713	0.701	0.6	0.598	0.569	0.589
	RPIfield	0.741	0.81	0.764	0.775	0.664	0.67	0.698	0.688
VGG - 19	Market-1501	0.716	0.721	0.741	0.713	0.645	0.632	0.634	0.628
	RAiD	0.803	0.848	0.799	0.818	0.7	0.685	0.693	0.695
	Partial-iLIDS	0.815	0.86	0.804	0.824	0.742	0.74	0.765	0.735
	RPIfield	0.777	0.768	0.785	0.773	0.672	0.67	0.641	0.661
ResNet - 18	Market-1501	0.813	0.882	0.836	0.847	0.736	0.742	0.77	0.76
	RAiD	0.788	0.793	0.813	0.785	0.717	0.704	0.706	0.7
	Partial-iLIDS	0.84	0.885	0.836	0.855	0.737	0.722	0.73	0.732
	RPIfield	0.852	0.897	0.841	0.861	0.779	0.777	0.802	0.772
ResNet - 34	Market-1501	0.814	0.805	0.822	0.81	0.709	0.707	0.678	0.698
	RAiD	0.85	0.919	0.873	0.884	0.773	0.779	0.807	0.797
	Partial-iLIDS	0.825	0.83	0.85	0.822	0.754	0.741	0.743	0.737
	RPIfield	0.83	0.875	0.819	0.839	0.757	0.755	0.78	0.75
ResNet - 50	Market-1501	0.842	0.873	0.8	0.828	0.687	0.685	0.656	0.676
	RAiD	0.828	0.897	0.851	0.862	0.751	0.757	0.785	0.775
	Partial-iLIDS	0.803	0.808	0.828	0.8	0.732	0.719	0.721	0.715
	RPIfield	0.867	0.912	0.863	0.882	0.764	0.749	0.757	0.759

has shown better performance when compared to other four models. Table IV records the performance of the models on benchmark datasets. All the models were trained from scratch on the same 8GB GPU with 16GB memory under similar initial conditions, except for the learning rate which has been selected differently to avoid overfitting. The stopping criteria is set as the flat validation error for more than 5 epochs and after two times decrease in learning rate.

The works in table IV are based on supervised and unsupervised methods that have clocked maximum mRA and CMC in the literature. We also compared with attention-based methods like AGW and the results show that the GAAP has indeed performed better than the AGW. The proposed GAAP has attention layers at the end of convolutional networks which enables the model to generate attention features for dense net classifier. However, the past attention-based methods used attention inside the convolutional layers that failed to capture the essential focused features for classification. We also found that the proposed model trains faster than the previous most popular triplet loss embedding with ResNet -50 backbone network. Fig. 8 and 9 shows the training accuracies and loss plots on Market – 1501 dataset for GAAP and DML with triplet loss respectively. Overall, our proposed GAAP model have shown good performance on RIPfiled Pe-reID dataset due to its rich multi view and multi resolution representation of the person images. Finally, the average recognition on all the benchmark datasets is around 84.12 which is 5% more than the previous methods.

V. Conclusion

In this work, we present an attention framework-based solution for person reidentification problem. The attention framework is built at the end of the feature extraction network and before the classifier dense network. Subsequently, the attention features are pooled using global averaging across the within class images. The proposed GAAP network is trained with ResNet – 50 as a backbone architecture for feature extraction. Consequently, extensive experimentation on four bench-

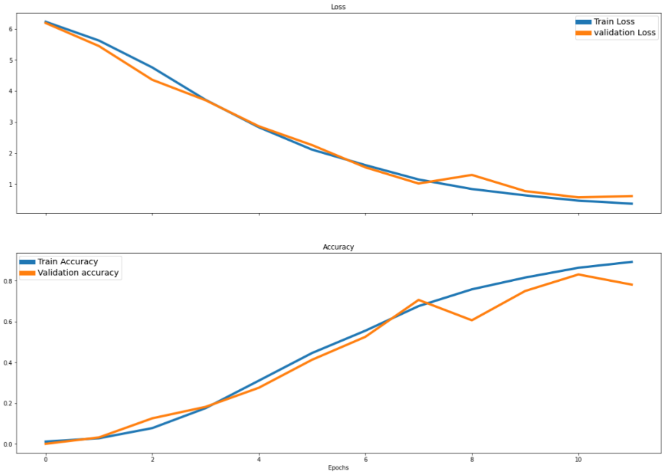


Fig. 8. Training Performance of GAAP with ResNet – 50 Backbone on Market – 1501 Dataset.

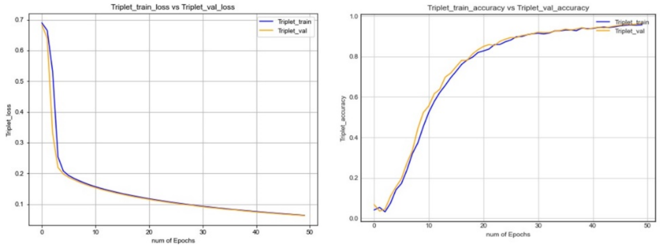


Fig. 9. Training Performance of DML with Triplet Loss Embedding with ResNet – 50 Backbone on Market – 1501 Dataset.

mark person Pe-reID datasets has shown that the proposed model performs better than state-of-the-art. Interestingly, the proposed model generated an average identification accuracy of around 84.12. Also, the proposed GAAP model trains in less time and achieves a competitive average validation accuracy on the benchmark datasets. However, the improvement of performance is achieved on large datasets with heterogeneous properties.

TABLE IV. Comparison of Previous State – of – the – Art Pe-reID Methods against the Proposed GAAP Model

Methods	Market - 1501		RAiD		Partial-iLIDS		RPIfield	
	mRA	CMC	mRA	CMC	mRA	CMC	mRA	CMC
PCB [28]	0.812	0.808	0.849	0.949	0.712	0.708	0.749	0.849
MGN [21]	0.832	0.821	0.852	0.952	0.732	0.721	0.752	0.852
HAN [3]	0.842	0.856	0.824	0.924	0.742	0.756	0.724	0.824
BDB [22]	0.803	0.817	0.785	0.885	0.703	0.717	0.685	0.785
IANet [19]	0.817	0.831	0.798	0.898	0.717	0.731	0.698	0.798
BoT [28]	0.824	0.838	0.812	0.912	0.724	0.738	0.712	0.812
AGW [2]	0.838	0.852	0.787	0.887	0.738	0.752	0.687	0.787
FPR [29]	0.823	0.837	0.798	0.898	0.723	0.737	0.698	0.798
PGFA [15]	0.813	0.827	0.843	0.943	0.713	0.727	0.743	0.843
HOReID [7]	0.824	0.838	0.891	0.991	0.724	0.738	0.791	0.891
PVPM [17]	0.822	0.836	0.802	0.902	0.722	0.736	0.702	0.802
GML [25]	0.813	0.827	0.813	0.913	0.713	0.727	0.713	0.813
HCT [30]	0.592	0.606	0.653	0.753	0.492	0.506	0.553	0.653
UDAML[24]	0.654	0.668	0.729	0.829	0.554	0.568	0.629	0.729
TLE [12]	0.713	0.727	0.758	0.858	0.613	0.627	0.658	0.758
MEB-Net [18]	0.752	0.766	0.765	0.865	0.652	0.666	0.665	0.765
PLF [16]	0.723	0.737	0.733	0.833	0.623	0.637	0.633	0.733
GAAP (OURS)	0.842	0.873	0.828	0.897	0.803	0.808	0.867	0.912

Acknowledgment

Author thanks the management of Koneru Lakshmiah Education Foundation and R&D department for funding the project.

References

[1] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1116–1124.

[2] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. Hoi, “Deep learning for person re-identification: A survey and outlook,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 6, pp. 2872–2893, 2021.

[3] L. Chen, H. Yang, Q. Xu, and Z. Gao, “Harmonious attention network for person re-identification via complementarity between groups and individuals,” *Neurocomputing*, vol. 453, pp. 766–776, 2021.

[4] D. Wu, S.-J. Zheng, X.-P. Zhang, C.-A. Yuan, F. Cheng, Y. Zhao, Y.-J. Lin, Z.-Q. Zhao, Y.-L. Jiang, and D.-S. Huang, “Deep learning-based methods for person re-identification: A comprehensive review,” *Neurocomputing*, vol. 337, pp. 354–371, 2019.

[5] G. Wang, J. Lai, P. Huang, and X. Xie, “Spatial-temporal person re-identification,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 8933–8940.

[6] Q. Leng, M. Ye, and Q. Tian, “A survey of open-world person re-identification,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1092–1108, 2019.

[7] G. Wang, S. Yang, H. Liu, Z. Wang, Y. Yang, S. Wang, G. Yu, E. Zhou, and J. Sun, “High-order information matters: Learning relation and topology for occluded person re-identification,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6449–6458.

[8] H.-X. Yu, A. Wu, and W.-S. Zheng, “Unsupervised person re-identification by deep asymmetric metric embedding,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 4, pp. 956–973, 2018.

[9] H. Zheng, X. Zhong, W. Huang, K. Jiang, W. Liu, and Z. Wang, “Visible-infrared person re-identification: A comprehensive survey and a new setting,” *Electronics*, vol. 11, no. 3, p. 454, 2022.

[10] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian, “Person re-identification in the wild,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1367–1376.

[11] D. Li, X. Chen, Z. Zhang, and K. Huang, “Learning deep context-aware features over body and latent parts for person re-identification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 384–393.

[12] Z. Tang and J. Huang, “Harmonious multi-branch network for person re-identification with harder triplet loss,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 18, no. 4, pp. 1–21, 2022.

[13] J. Yu and H. Oh, “Graph-structure based multi-label prediction and classification for unsupervised person re-identification,” *Applied Intelligence*, pp. 1–13, 2022.

[14] Y. Li, L. Liu, L. Zhu, and H. Zhang, “Person re-identification based on multi-scale feature learning,” *Knowledge-Based Systems*, vol. 228, p. 107281, 2021.

[15] J. Miao, Y. Wu, P. Liu, Y. Ding, and Y. Yang, “Pose-guided feature alignment for occluded person re-identification,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 542–551.

[16] Y. Sun, Q. Xu, Y. Li, C. Zhang, Y. Li, S. Wang, and J. Sun, “Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 393–402.

[17] S. Gao, J. Wang, H. Lu, and Z. Liu, “Pose-guided visible part matching for occluded person reid,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 744–11 752.

[18] Y. Zhai, Q. Ye, S. Lu, M. Jia, R. Ji, and Y. Tian, "Multiple expert brainstorming for domain adaptive person re-identification," in *European Conference on Computer Vision*. Springer, 2020, pp. 594–611.

[19] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan, and X. Chen, "Interaction-and-aggregation network for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9317–9326.

[20] Z. Wang, J. Jiang, Y. Wu, M. Ye, X. Bai, and S. Satoh, "Learning sparse and identity-preserved hidden attributes for person re-identification," *IEEE Transactions on Image Processing*, vol. 29, pp. 2013–2025, 2019.

[21] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 274–282.

[22] Z. Dai, M. Chen, X. Gu, S. Zhu, and P. Tan, "Batch dropblock network for person re-identification and beyond," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 3691–3701.

[23] H. Kim, H. Kim, B. Ko, J. Shim, and E. Hwang, "Two-stage person re-identification scheme using cross-input neighborhood differences," *The Journal of Supercomputing*, vol. 78, no. 3, pp. 3356–3373, 2022.

[24] R. Pierre and M. Qi, "Unsupervised domain adaption based on metric learning for person re-identification," in *2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication (CTISC)*. IEEE, 2021, pp. 417–421.

[25] J. Meng, W.-S. Zheng, J.-H. Lai, and L. Wang, "Deep graph metric learning for weakly supervised person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[26] A. Chakraborty, A. Das, and A. K. Roy-Chowdhury, "Network consistent data association," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 9, pp. 1859–1871, 2015.

[27] M. Zheng, S. Karanam, and R. J. Radke, "Rpifield: A new dataset for temporally evaluating person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 1893–1895.

[28] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person re-identification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.

[29] L. He, Y. Wang, W. Liu, H. Zhao, Z. Sun, and J. Feng, "Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 8450–8459.

[30] K. Zeng, M. Ning, Y. Wang, and Y. Guo, "Hierarchical clustering with hard-batch triplet loss for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13 657–13 665.

Towards a Richer IndoWordNet with New Additions for Hindi and Gujarati Languages

Milind Kumar Audichya[1]
SJD International,
Surat, India - 395009

Jatinderkumar R. Saini[2]*
Symbiosis Institute of Computer Studies
and Research, Symbiosis International
(Deemed University), Pune, India - 411016

Jatin C. Modh[3]
Gujarat Technological University,
Ahmedabad, India - 382424

***Abstract*—The authors of this research paper present a mechanism for dealing with loanwords, missing words, and newly developed terms inclusion issues in WordNets. WordNet has evolved as one of the most prominent Natural Language Processing (NLP) toolkits. This mechanism can be used to improve the WordNet of any language. The authors chose to work with the Hindi and Gujarati languages in this research work to achieve a higher quality research aspect because these are the languages with major dialects. The research work used more than 5000 Hindi verse-based data corpus instead of a prose-based data corpus.As a result, nearly 14000 Hindi words were discovered that were not present in the popular Hindi IndoWordNet, accounting for 13.23 percent of the total existing word count of 105000+. Working with idioms was a distinct method for the Gujarati language. Around 3500 idioms data were used, and nearly 900 Gujarati terms were discovered that did not exist in the IndoWordNet, accounting for nearly 1.4 percent of the total of 64000+ Gujarati words in the IndoWordNet. It will also contribute almost 14000 Hindi words and around 900 Gujarati words to the IndoWordNet project.**

***Keywords*—Gujarati; Hindi; Indian language WordNet; IndoWordNet; loanwords; WordNet**

I. Introduction

Languages transmit knowledge from one generation to the next, as well as from one culture to another, through communication. There are multiple modes of communication, including speech, writing, and sign language. No matter what the communication medium is, Communication becomes a systematic process when using any language because each has its vocabulary, grammar, and components. There are 7151 different languages in the world [1], some of which are well-known while others are on the verge of extinction.

Hindi [2], a truly mellifluous Indo-Aryan language, is one of the most popular languages in the world, which is scripted using Devanagari [3] and is currently supported by The Unicode Standard [4]. Although Hindi is widely spoken throughout the world, it is mostly used in India, particularly in the Hindi Belt, which encompasses sections of India's four major zones: eastern, western, central, and northern [2]. The rich literature and long history of Hindi users make up the language's legacy. Hindi-based research and related efforts in relevant research disciplines are currently strongly emphasized.

Gujarati[5], like Hindi, is a sweet-sounding language that belongs to the Indo-Aryan language family. Although Gujarati is used by Gujaratis all over the world, it is especially utilised for communication in Gujarat, located in India's western region, which is considered the origin place of Gujarati. Gujarati is written in the Devanagari script as well and is currently supported by The Unicode Standard [6]. Gujarati is also thriving in terms of research and development in recent years [7], [8].

Natural languages are those that have evolved gradually and are used by humans to communicate. It is a well-known fact that computers don't understand natural languages. To make computers understand different kinds of languages, many research works are going on for different languages. Natural Language Processing (NLP) [9] is used to assist computers in understanding these languages. Computer Science (CS) [10], Computational Linguistics (CL) [11], and Artificial Intelligence (AI) [12] all intertwine in NLP. To grasp a language, computers, like humans, must understand the alphabet, words, meanings of words, pronunciation, vocabulary, sentence structure, context, and all grammatical rules associated with it.

Because computers lack cognitive intelligence, making them grasp any language is difficult. WordNet [13]–[15], a systematically managed correlated lexical database that usually consists of words and semantic relations with the words including synonyms, hyponyms, and meronyms, is commonly used to help computers overcome this limitation. As a practical matter, WordNet can be thought of as a hybrid of a dictionary and a thesaurus. There are various WordNet-based research projects underway [16], [17], some of which are language-specific exclusively and others that are multilingual.

The authors used IndoWordNet [18], [19], a well-known WordNet based on Indian regional languages, for this study. IndoWordNet is a WordNet that has a base of 18 different Indian languages. The Center for Indian Language Technology (CFILT) in the Computer Science and Engineering Department at IIT Bombay created IndoWordNet. Hindi is the default base language of this WordNet. The IndoWordNet is used in this research for both Hindi and Gujarati languages. Even when a well-built WordNet exists for popular languages [20], [21], there is always room for improvement.

The authors were motivated by the various WordNets they used and the difficulties they encountered while using them in different research studies due to missing or borrowed words. Current research is an attempt to address such concerns. This research will be useful in strengthening the WordNet and adding new words, loan words, and missing words.

Loanwords, missing words and newly developed terms are

*Corresponding authors.

several issues that must be addressed in the building of any language's WordNet. Each language typically has loanwords, which are words taken from other languages with slight or no alterations. Another challenge is dealing with words that are not in WordNet and dealing with the newly developed terms in the recent times. This research aims to efficiently address these issues in order to improve WordNets throughout time.

This research paper will also explain how to use WordNets to make effective use of information created by WordNets in order to strengthen WordNet research. The Section II, Literature Review will discuss various research works in a similar or related field to determine the research gap and the need for current research work. The Section III, Methodology segment describes the actual mechanism that can be used to continuously improve the WordNets. The Section IV, Results part will summarise the outcome of the applied mechanism for the Hindi and Gujarati languages utilising the IndoWordNet. The Section V, Conclusion portion will indicate how the improvements should be implemented and expanded. Finally, Section VI, Future Enhancement include some relevant and helpful remarks for upcoming research works.

II. Literature Review

The authors attempted to delve deeper into the WordNet research area first, as they were working with various WordNets and encountering issues with missing words from WordNets. WordNets are lexical hybrid resources that are an interconnected combination of a dictionary and a thesaurus built using different approaches for different languages. Professor George A. Miller[13], [14] oversaw the creation of the first WordNet for the English language at Princeton University's cognitive science lab. Following the development of Princeton's American English WordNet, a multilingual wordnet EuroWordNet[15] was created on similar principles during March 1996 to June 1999, specifically designed for the various European languages (Dutch, Italian, Spanish, German, French, Czech and Estonian). The development of WordNets for various Indian languages began in the early 2000s, spearheaded by the Hindi WordNet [16]. Later, in the direction of Bhattacharyya research project IndoWordNet [17] was expanded to include multiple Indian languages, and it now supports 18 languages (Assamese, Bangla, Bodo, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Tamil, Telugu, Urdu).

Authors attempted to locate some advancement and improvement associated research works related to WordNet after exploring for foundation-related research works. Redkar et al. [18] attempted to create an online multilingual dictionary with 19 languages for researchers and non-researchers to utilise for various purposes. To access the IndoWordNet [17], Panjwani et al. created pyiwn [19], a python-based Application Programming Interface (API), pyiwn is used to access IndoWordNet in this study.

In other recent work, McCrae et al. [20] worked on English WordNet 2020, an open-source project to improve and extend the Princeton WordNet, which hasn't been updated in a long time. In other recent work, McCrae et al. worked on English WordNet 2020, an open-source project to improve and extend the Princeton WordNet, which hasn't been updated in a long time. They made around 15000 modifications since the last version was updated.

Kanojia et al. [21] worked on linking 18 different Indian Language WordNets to Princeton WordNet, which aids in the exchange of knowledge and comprehension of various terminologies and their meanings in a multilingual environment. Fellbaum [22] described the WordNet's latest approaches, which are focused on language mapping concepts. He also discusses Crosslinguistic WordNets, as well as all of the components of how WordNets are managed over time.

In an another introductory work, Bhensdadia et al. [23] reviewed the development of the Gujarati wordnet. That research was also a component of the IndoWordNet, and the source language for Gujarati language development was solely Hindi. As a result, the authors of the current study decided to conduct parallel research in both languages. Zankhana and Sajja [24] also worked on the issue of word sense disambiguation in the Gujarati language using a Knowledge-Based Approach and a Genetic Algorithm. Using the IndoWordNet, Modh and Saini attempted to contextually improve Gujarati machine translation [25]. They used the n-gram model and tests with varying frame sizes to try to enhance the translation of Gujarati idioms.

The authors reviewed numerous study works linked to WordNets of various languages, as well as improvement-related works and a few studies focusing on the trip of the WordNets' development thus far.The authors realised that there is still a need to identify some systematic technique that might aid streamline the process of improving and strengthening WordNets. Inclusion of loanwords, missing words, and newly formed terminology in any WordNet of any language around the world must be done in a systematic and organized way.

The authors also discovered that while most WordNet-based initiatives focus on prose-based content, verse-based content can also provide some noteworthy contributions. Using verse for such research purposes is still a challenge in and of itself. Languages with a greater number of dialects make such study even more difficult. Taking into account all of these factors, the authors decided to focus on developing a system for continual WordNet improvement and strengthening. The authors opted to work with Hindi and Gujarati languages since they have diverse dialects, and they chose to evaluate the mechanism's quality and efficacy with Hindi's verse-based literary content and Gujarati's idiom-based data.

Following an extensive literature review and identifying a pressing need of the hour while conducting research on metadata generation for Hindi poetry [26], [27] and a few other research works related to Hindi poetry [28] and Gujarati idioms [29] using WordNet, the findings indicate the need for a systematic approach to dealing with loanwords, missing words, and recently developed terminologies. Since there are so many dialects in Hindi and Gujarati, resolving such difficulties is more complicated; Intriguingly, the authors acknowledged this research as a challenging one.

III. Methodology

This proposed research revolves around the use of WordNet, specifically the use of WordNet to find words that aren't currently in the WordNet. Several phases of this research as represented in Fig. 1, Core Methodology includes:

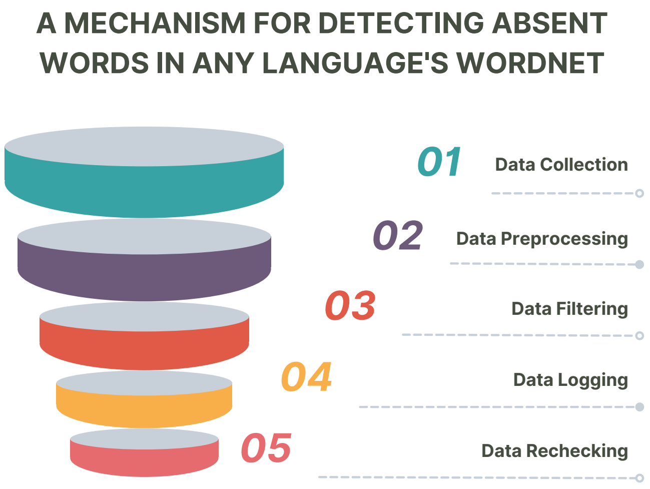


Fig. 1. Core Methodology

- III-A Data Collection
- III-B Data Preprocessing
- III-C Data Filtering
- III-D Data Logging
- III-E Data Rechecking

Authors incorporated transliteration and translation wherever non-English text is included, keeping in mind the international readership of the research work.

A. Data Collection

Special efforts were expended during the data collection phase for both Hindi and Gujarati. After analysing and brainstorming numerous types of data, the Hindi language poetic data and the Gujarati language idiom data were chosen with the goal of finding the most missing terms for a qualitative and quantitative contribution to the improvement of the respective languages' WordNets. Online websites and portals, as well as offline literature, were thoroughly examined for data collection. This study makes use of poetic data from Hindi literature. In Hindi poetry, verses play an important role. To write any type of verse, there are some specific rules that must be followed. Audichya and Saini's Hindi Verse dataset [30], as well as some additional data, were used. A total of 5011 poetic pieces of information were collected.

Example Hindi Data:

लग रहा है आज अम्बर, ज्यों उमड़ता सा समन्दर।
उठ चली चंचल हिलोंरे, आ रहे जैसे बवंडर।।

Transliteration of Example Hindi Data:

Lag rahā hai āj ambara, jyoan umadatā sā samandara.
uṭh chalī chanchal hilore, ā rahe jaise bavanḍara..

Translation of Example Hindi Data:

Looks like the sky today, as a rising sea.
Agitated tremors got up, like a tornado coming.

Gujarati data was obtained in the same way that Hindi data was, from various books and portals, along with some data which was directly acquired from the authors' previous research studies [31]. All of the standard Unicode-based data was collected, organised, and verified by specialists between December 2017 and April 2022.

Example Gujarati Data:

મોઢા પર મારવું

Transliteration of Example Gujarati Data:

Moḍhā par māravun

Translation of Example Gujarati Data:

Slap on face

B. Data Preprocessing

Following data collection, data processing takes place, which includes data preprocessing and cleaning. With the exception of the specific language's unicode range, which eventually removes unnecessary special symbols, emojis, junk characters, and so on, all unnecessary data is discarded first. Punctuation marks are also removed (Full Stop, Comma, Question Mark, Colan, Semicolon, Brackets, Exclamation Mark, Quotation Mark, Slash, and all irrelevant marks). The data is chunked into a list of words after the marks are removed and the preprocessing cleaning operation is completed. This procedure is known as Word Tokenization in the domain of NLP.

Preprocessed Hindi Data:

लग रहा है आज अम्बर ज्यों उमड़ता सा समन्दर उठ
चली चंचल हिलोंरे आ रहे जैसे बवंडर

Transliteration of Preprocessed Hindi Data:

Lag rahā hai āj ambara jyoan umadatā sā
samandara uṭh chalī chanchal hilore ā
rahe jaise bavanḍara

Translation of Preprocessed Hindi Data:

Looks dwell is today sky like surge as
sea rise walk fickle tremors come stay like
tornado

Preprocessed Gujarati Data:

મોઢા પર મારવું

Transliteration of Preprocessed Gujarati Data:

Moḍhā par māravun

Translation of Preprocessed Gujarati Data:

Slap on face

C. Data Filtering

The data preprocessing stage's list of words is now ready to be filtered. Let us first define filtering and its purpose. Every language has its own StopWords. Stopwords are commonly used words that have a little or no impact on the meaning of a sentence. Because this experiment uses Hindi language data, the authors used Hindi StopWords from a hybrid StopWords-based research of Hindi [32] and Gujarati[33] languages. StopWords from a specific language must be used in this stage to filter based on the language chosen for the implementation of this mechanism. StopWords were filtered while applying filtering to the list of words produced by the data processing stage.

Found Hindi Stop Words:

रहा है आज रहे जैसे

Transliteration of Hindi Stop Words:

rahā hai āj rahe jaise

Translation of Hindi Stop Words:

dwell is today stay like

Filtered Hindi Data:

लग अम्बर ज्यों उमड़ता सा समन्दर उठ चली चंचल

हिलोरे आ बवंडर

Transliteration of Filtered Hindi Data:

Lag ambara jyoan umadatā sā samandara

uṭh chalī chanchal hilore ā bavanḍara

Translation of Filtered Hindi Data:

Looks sky like surge as sea rise walk

fickle tremors come tornado

For the given Hindi example, five Hindi StopWords were filtered.

Found Gujarati Stop Word:

પર

Transliteration of Found Gujarati Stop Word:

par

Translation of Found Gujarati Stop Word:

on

Filtered Gujarati Data:

મોઢા મારવું

Transliteration of Filtered Gujarati Data:

Moḍhā māravun

Translation of Filtered Gujarati Data:

Slap face

For the given Gujarati example, a Gujarati StopWord was filtered. Data filtering is critical because it will assist us in reducing computational processing in subsequent stages. The remaining words will now go through the logging process.

D. Data Logging

The act of keeping a record of something is referred to as logging. Words from the filtered data list are checked in WordNet one by one at this stage. If a word is already in the WordNet, one can access all of the relevant properties of that word, such as Synsets, Synonyms, POS tags, Gloss, Example statements, and much more relevant information provided by the specific WordNet.

Now comes the crucial part of this research work: if any of the words are not found while checking the word's existence in the WordNet, some WordNets may produce an information message, while others may generate an error through exceptions. Such cases must be handled properly, and any words that are not found in WordNet must be logged and kept in a list of not found words.

The following words were not found in WordNet for the filter data from the data filtering stage.

Logged Hindi Data:

लग ज्यों उमड़ता सा उठ चली चंचल हिलोरे

Transliteration of Logged Hindi Data:

Lag jyoan umadatā sā uṭh chalī chanchal

hilore

Translation of Logged Hindi Data:

Looks like surge as rise walk fickle tremors

Logged Gujarati Data:

મોઢા

Transliteration of Logged Gujarati Data:

Moḍhā

Translation of Logged Gujarati Data:

face

This is a continuous process that can be repeated whenever a word is not found in WordNet. This entire mechanism can be embedded in any system that uses WordNet. All that remains is to integrate these various stages and keep track of words that are not found in WordNet while searching for various types of uses.

E. Data Rechecking

This is an optional step in reprocessing the logged data by rechecking it against WordNet. If this mechanism is integrated with any other system that uses WordNet, the logging stage will produce a large list of words that were not found over time.

Now, that list may contain some words that have already been added to WordNet in a later release, so to avoid those words from the logged data, the entire data can be rechecked with the most recent updated WordNet, and any duplicate words that have already been added to WordNet will be removed. Similarly, logged data can be filtered again with the updated StopWords if necessary.

The logged Hindi and Gujarati data example has been rechecked, and one word has been added in Hindi WordNet in the most recent update, so it has been removed from the data. Gujarati data remains as it is as the word is still not added in the Gujarati WordNet.

Rechecked Hindi Data:

लग ज्यों उमड़ता सा उठ चली हिलोरे

Transliteration of Logged Hindi Data:

Lag jyoan umadatā sā uṭh chalī hilore

Translation of Logged Hindi Data:

Looks like surge as rise walk tremors

Rechecked Gujarati Data:

મોઢા

Transliteration of Rechecked Gujarati Data:

Moḍhā

Translation of Rechecked Gujarati Data:

face

As a result, the data rechecking stage is for final checks before producing a clean list of data that are not available in WordNet. Let's take a look at the overall results of this study.

F. Discussion

If this mechanism is applied and followed in a systematic manner, this research will undoubtedly help to strengthen the various WordNet-based research works. This mechanism is useful for almost all WordNet-based projects, regardless of language. It was purposefully tested with Hindi language based poetic data and Gujarati idiom based data to determine its effectiveness because processing poems and idioms differs slightly from processing prose. That is the only explanation for such extraordinary results. There could be several reasons for the large number of words that were not found. It is possible to mention a few borrowed words, missing words, newly developed terminology, and combined or misspelt words. Because Hindi and Gujarati have so many dialects, there are more chances of borrowing words from neighbouring and sister languages.

IV. Results

This research was carried out over a long period of time, between December 2017 and April 2022. During this time, 5011 Unicode Hindi Verse-based poetic literature data and 3472 Gujarati idioms data were processed using this research methodology through all of the stages III-A. Data Collection, III-B. Data Preprocessing, III-C. Data Filtering, III-D. Data Logging, III-E. Data Rechecking as described in the III. Methodology section. Table I. Overall Results representing the different stats of current research work. As a result, the authors were able to populate a list [34] of 13,593 Hindi words which are not available in the Hindi WordNet section of Indian Languages WordNet (IndoWordNet). The total number of Hindi words in the IndoWordNet is 1,05,458, but through this research work, 13,953 new potential words were discovered, accounting for nearly 13.23 percent of the total number of words. In addition, 887 Gujarati words[35] were discovered while analysing 3,472 Gujarati idioms, accounting for 1.38 percent of the existing 64,300 Gujarati words in WordNet. This many words are more than enough to demonstrate the utility of this mechanism. Because there haven't been any similar research and datasets based on the Hindi and Gujarati languages, benchmarking isn't possible at this moment. Due to the limitation that there are currently no other comparable datasets available for both Hindi and Gujarati, future results may vary depending on the availability. If used, the current research methodology can yield significant results for various languages as well. This mechanism will undoubtedly aid in the improvement of WordNets in various languages around the world. While using this mechanism with other languages, the results may vary, but it will significantly improve any WordNet.

TABLE I. Overall Results

Sr. No.	**1**	**2**
Language	Hindi	Gujarati
Existing Words in IndoWordNet	105458	64300
Absent Words in IndoWordNet	13953	887
Absent Words %	13.23%	1.38%

V. Conclusion

To summarise, continual efforts are always required to strengthen the WordNets in order to accommodate freshly produced terms, loanwords, and missing words. Dealing with WordNets of languages with several dialects is difficult.Maintaining a log of words not found in WordNet while using any WordNet is usually beneficial. Later, processing those words with all of the WordNet's accessible words will yield a potential list of words. Such lists may be evaluated for inclusion in WordNet. On such words, the regular techniques for adding words to any language's WordNet can be used.

If these words are still not related to the specific language, they may have been borrowed from another language. In that scenario, these words can be cross-checked with WordNets from neighbouring languages from which they may have borrowed. For the continuing enhancement and strengthening of WordNet research projects, such methods and mechanisms can be included with practically every language's WordNet.

VI. Future Enhancement

The list of populated words after the data rechecking phase can be checked with the nearby and sister languages WordNets to identify the words in case some of the borrowed words belong to some nearby languages for future enhancement. Other studies, such as joint words or language detection related research, studies can also be conducted to check for missspelt and incorrect words.

References

[1] Ethnologue.com How many languages are there in the world?. *Ethnologue.com*. (2022), https://www.ethnologue.com/guides/how-many-languages

[2] Contributors, W. Hindi - Wikipedia. *Wikipedia.org*. (2021), https://en.wikipedia.org/wiki/Hindi

[3] Contributors, W. Devanagari - Wikipedia. *Wikipedia.org*. (2021), https://en.wikipedia.org/wiki/Devanagari

[4] Unicode, I. Devanagari - The Unicode Standard. *The Unicode Standard,*. **14** (2021), https://unicode.org/charts/PDF/U0900.pdf

[5] Contributors, W. Gujarati - Wikipedia. *Wikipedia.org*. (2021), https://en.wikipedia.org/wiki/Gujarati_language

[6] Unicode, I. Gujarati - The Unicode Standard. *The Unicode Standard,*. **14** (2021), https://unicode.org/charts/PDF/U0A80.pdf

[7] Audichya, M.K. & Saini, J.R. A Study to Recognize Printed Gujarati Characters Using Tesseract OCR. *Engineering, Technology And Applied Science Research*. **5** pp. 1505-1510 (2017,9)

[8] Modh, J.C. & Saini, J.R. Context Based MTS for Translating Gujarati Trigram and Bigram Idioms to English. *2020 International Conference For Emerging Technology (INCET)*. pp. 1-6 (2020), https://doi.org/10.1109/INCET49848.2020.9154112

[9] Contributors, W. Natural Language Processing - Wikipedia. *Wikipedia.org*. (2022), https://en.wikipedia.org/wiki/Natural_language_processing

[10] Contributors, W. Computer Science - Wikipedia. *Wikipedia.org*. (2022), https://en.wikipedia.org/wiki/Computer_science

[11] Contributors, W. Computational Linguistics - Wikipedia. *Wikipedia.org*. (2022), https://en.wikipedia.org/wiki/Computational_linguistics

[12] Contributors, W. Artificial Intelligence - Wikipedia. *Wikipedia.org*. (2022), https://en.wikipedia.org/wiki/Artificial_intelligence

[13] Miller, G. WordNet: A Lexical Database for English. *Commun. ACM*. **38**, 39-41 (1995,11), https://doi.org/10.1145/219717.219748

[14] Miller, G. WordNet: An electronic lexical database. (MIT press,1998)

[15] Vossen, P. EuroWordNet: A multilingual database with lexical semantic networks. *Dordrecht: Kluwer Academic Publishers.*. **10** (1998), https://link.springer.com/book/10.1007/978-94-017-1491-4

[16] Narayan, D., Chakrabarti, D., Pande, P. & Bhattacharyya, P. An experience in building the indo wordnet-a wordnet for hindi. *First International Conference On Global WordNet, Mysore, India*. **24** (2002), https://www.academia.edu/314054/

[17] Bhattacharyya, P. IndoWordNet. *Proceedings Of The Seventh International Conference On Language Resources And Evaluation (LREC'10)*. (2010,5), http://www.lrec-conf.org/proceedings/lrec2010/pdf/939_Paper.pdf

[18] Redkar, H., Singh, S., Joshi, N., Ghosh, A. & Bhattacharyya, P. IndoWordNet Dictionary: An Online Multilingual Dictionary using IndoWordNet. *Proceedings Of The 12th International Conference On Natural Language Processing*. pp. 71-78 (2015,12), https://aclanthology.org/W15-5910

[19] Panjwani, R., Kanojia, D. & Bhattacharyya, P. pyiwn: A Python based API to access Indian Language WordNets. *Proceedings Of The 9th Global Wordnet Conference*. pp. 378-383 (2018,1), https://aclanthology.org/2018.gwc-1.47

[20] McCrae, J., Rademaker, A., Rudnicka, E. & Bond, F. English WordNet 2020: Improving and Extending a WordNet for English using an Open-Source Methodology. *Proceedings Of The LREC 2020 Workshop On Multimodal Wordnets (MMW2020)*. pp. 14-19 (2020,5), https://aclanthology.org/2020.mmw-1.3

[21] Kanojia, D., Patel, K. & Bhattacharyya, P. Indian Language Wordnets and their Linkages with Princeton WordNet. (arXiv,2022), https://doi.org/10.48550/arxiv.2201.02977

[22] Fellbaum, C. WordNet. *Theory And Applications Of Ontology: Computer Applications*. pp. 231-243 (2010), https://doi.org/10.1007/978-90-481-8847-5_10

[23] Bhensdadia, C., Bhatt, B. & Bhattacharyya, P. Introduction to Gujarati wordnet. *Third National Workshop On Indowordnet Proceedings*. **494** (2010)

[24] Vaishnav, Z. & Sajja, P. Knowledge-based approach for word sense disambiguation using genetic algorithm for Gujarati. *Smart Innovation, Systems And Technologies*. **106** pp. 485-494 (2019), http://dx.doi.org/10.1007/978-981-13-1742-2_48

[25] Modh, J. & Saini, J. Using IndoWordNet for Contextually Improved Machine Translation of Gujarati Idioms. *International Journal Of Advanced Computer Science And Applications (IJACSA)*. **12**, pp. 225-232 (2021), http://dx.doi.org/10.14569/IJACSA.2021.0120128

[26] Audichya, M.K. & Saini, J.R. Computational linguistic prosody rule-based unified technique for automatic metadata generation for Hindi poetry. *2019 1st International Conference On Advances In Information Technology (ICAIT)*. pp. 436-442 (2019,7), https://ieeexplore.ieee.org/document/8987239/

[27] Audichya, M.K. & Saini, J.R. Stanza Type Identification using Systematization of Versification System of Hindi Poetry. *International Journal Of Advanced Computer Science And Applications*. **12**, pp. 142-153 (2021), https://dx.doi.org/10.14569/IJACSA.2021.0120117

[28] Audichya, M.K. & Saini, J.R. Towards Natural Language Processing with Figures of Speech in Hindi Poetry. *International Journal Of Advanced Computer Science And Applications*. **12**, pp. 128-133 (2021), https://dx.doi.org/10.14569/IJACSA.2021.0120316

[29] Modh, J.C. & Saini, J.R. Using IndoWordNet for Contextually Improved Machine Translation of Gujarati Idioms. *International Journal Of Advanced Computer Science And Applications*. **12**, pp. 225-232 (2021), http://dx.doi.org/10.14569/IJACSA.2021.0120128

[30] Audichya, M.K. & Saini, J. R. Hindi Verse Dataset. *Mendeley Data*. (2022), https://data.mendeley.com/datasets/cp6htsbbpp/1

[31] Saini, J.R. & Modh, J.C. GIdTra: A dictionary-based MTS for translating Gujarati bigram idioms to English. *2016 Fourth International Conference On Parallel, Distributed And Grid Computing (PDGC)*. pp. 192-196 (2016)

[32] Jha, V., Manjunath, N., Deepa Shenoy, P. & R, V. Hindi Language Stop Words List. *Mendeley Data*. **V1** (2018), https://data.mendeley.com/datasets/bsr3frvvjc/1

[33] Quanteda Initiative. A List 210 Gujarati Stop Words. *Github.com*. (2022), https://github.com/quanteda/stopwords/issues/11

[34] Audichya, M.K. & Saini, J.R. Additional Hindi Words for IndoWordNet. *Mendeley Data*. (2022), https://data.mendeley.com/datasets/db8sh8js67/1

[35] Audichya, M.K., Saini, J.R. & Modh J.C. Additional Gujarati Words for IndoWordNet. *Mendeley Data*. (2022), https://data.mendeley.com/datasets/3jtm7htsyt/1

Prediction of Diabetic Retinopathy using Convolutional Neural Networks

Manal Alsuwat, Hana Alalawi, Shema Alhazmi, Sarah Al-Shareef
Computer Science Department, College of Computer Science and Information System
Umm AlQura University, Makkah, 21955, Saudi Arabia

Abstract—**Diabetic retinopathy (DR) is among the most dangerous diabetic complications that can lead to lifelong blindness if left untreated. One of the essential difficulties in DR is early discovery, which is crucial for therapy progress. The accurate diagnosis of the DR stage is famously complicated and demands a skilled analysis by the expert being of fundus images. This paper detects DR and classifies its stage using retina images by applying conventional neural networks and transfer learning models. Three deep learning models were investigated: trained from scratch CNN and pre-trained InceptionV3 and EfficientNetsB5. Experiment results show that the proposed CNN model outperformed the pre-trained models with a 9 to 25% relative improvement in F1-score compared to pre-trained InceptionV3 and EfficientNetsB5, respectively.**

Keywords*—*CNN; convolutional neural networks; deep learning; transfer learning; medical imaging; diabetic retinopathy; retina fundus images

I. Introduction

Diabetic retinopathy (DR) is one of the diseases associated with diabetes and causes blindness to 4.4 million Americans over age 40 [1]. DR is an eye condition developed quickly in diabetes mellitus patients in type 1 or 2 [2]. DR often has no obvious symptoms in the early stages, but it becomes more pronounced as the disease progresses to more severe stages. An experienced ophthalmologist schedules a plan, which may run from weeks to months, to examine diabetic patients to determine their stage based on the retina's lesions and the severity level. Essentially, DR affects light-sensitive tissue blood vessels (i.e., retina) [3]. DR can be either nonproliferative or proliferative. In nonproliferative DR (NPDR), no abnormal blood vessel growth is found in the retina. Still, small outpouchings exist as the wall of retinal capillaries is weakened due to high blood glucose. These outpouchings are known as microaneurysms. NPDR can be mild, moderate, or severe based on the number of found microaneurysms and distortion of the blood vessels in the retinal exam. As the disease progresses, blood vessels may grow abnormally covering the retina; hence, DR becomes proliferative (PDR), leading to severe visual consequences.

In preventing blindness caused by the DR, detection, diagnosis, and treatment in earlier stages will control the disease and reduce vision loss. Diagnosis of the DR is complicated and requires high potential abilities [4]. One well-known obstacle for DR is that even for diabetic macular oedema, there are no early warning signs. Therefore, it is highly desirable to detect DR on time. Currently, DR diagnosis needs a well-trained doctor to diagnose the disease and manually evaluate digital images of the fundus of the retina. DR is recognised through identifying lesions connected with vascular malformations resulting from diabetes. This process may require longer time and effort depending on the experience and efficiency of the examiner doctor.

With the recent advancements in intelligent solutions, deep learning and transfer learning techniques showed significant success in object recognition and detection tasks. This research aims to automate DR diagnosis through exploiting convolutional neural networks (CNN) and transfer learning to identify DR from retina images. The Asia Pacific Tele-Ophthalmology Society (APTOS) dataset was used for blindness diagnosis and detection in this research. In addition, a comparison of the evaluation of different models to detect the disease effectively. This intelligent solution would help the health community diagnose the disease more efficiently, using less time and resources.

The remainder of this paper is organised as follows. In Section II, the related studies on the topic have been reviewed. Section III presents the research models used in this study followed by a description of the dataset used in the diagnosis DR in Section IV. Section V lays evaluation metrics and experimental design. Section VI presents the results of the experiments, and finally, the study is concluded in Section VII.

II. Related Work

Early DR detection is critical and time-consuming, and ophthalmologists are burdened. This attracted many researchers to develop early DR detectors and classifiers. Here, an overview of the deep learning techniques used in the previous literature is presented. Also, the used DR datasets in those studies are summarised. All of the reviewed literature detected DR from retinal fundus images. If detected, DR was classified into one of four severity levels: mild, moderate, severe NPDRs and PDR.

In the deep learning approach, CNN extracts features from input images and feeds them to the deeper layer in the model. Shan et al. [19] distinguished microaneurysms from fundus images using stacked sparse autoencoder (SSAE). Their model reached 91.3% for F1-score and an AUC of 96.2%. Singh et al. [20] employed a densely connected neural network architecture to detect the DR severity efficiently. Experimental findings showed that the DR severity could be successfully identified through the model with an accuracy of 83.6%.

Some researchers fine-tuned pre-trained models, known as transfer learning (TL), instead of training their models from scratch. These pre-trained models were initially trained using a large amount of out-of-domain data for object recognition and

TABLE I. Previous Work in Early DR Detection using Transfer Learning. Preprocessing Column Indicates whether any Image Processing was Applied before Model Fine-Tuning.

Work	Preprocessing	Pre-trained models	Dataset	Performance
Nquyen et al. [5]	Highlight spot, Crop, Drop outliers, Convert B/W, Rotate tree, Special Filtering	VGG-16, VGG-19	EyePACS	(Ensemble) Accuracy 82% Sensitivity 80% Specificity 82%
Masood et al. [6]	Downsize to a common radius, Normalize, Crop a borders	InceptionV3	EyePACS	Accuracy 48.2%
Maswood et al. [7]	Ben's preprocessing	EfficientNet-B5	APTOS2019	Accuracy 93%
AbdelMaksoud et al. [8]	Filtering using median filter, Resize into 256 × 256, Transformation Processes, Normalize	EffecientNet-B0	IDRiD	Accuracy 86%
Qummar et al. [9]	Resize into 786 × 512, Mean normalized	Resnet50, Inceptionv3, Xception,Dense121, Dense169	EyePACS	(Ensemble) Accuracy 80.8% Sensitivity 51.5% Specificity 86.7%
Gao et al. [10]	Remove black borders, Resize into 672 × 672	(modified) MobileNet-Dense, MobileNetV2	MESSIDOR, EyePACS	(Ensemble) Accuracy 96.2%
Taufiqurrahman et al. [11]	Resize 224 × 224	MobileNetV2-SVM, MobileNetV2	APTOS2019	(MobileNetV2-SVM) Accuracy 85%
Khojasteh et al. [12]	Patch Exraction	SVM/KNN/OPF, DRBM,ResNet-50	DIARETDB1, e-Ophtha	(ResNet-50-SVM) Accuracy 98.2% Sensitivity 99% Specificity 96%
Hemanth et al. [13]	Histogram equalisation	—	MESSIDOR	Accuracy 97% Sensitivity 94% Specificity 98%
Kose et al. [14]	Kirsch's template	—	MESSIDOR	Accuracy 89.8% Sensitivity 79.6% Specificity 93.2%
Pham et al. [15]	Subtracting the average local colour using a Gaussian mask	EfficientNet-B5	APTOS2019	—
Shankar et al. [16]	Histogram based segmentation	ResNet50	MESSIDOR	Accuracy 99.28% Sensitivity 98% Specificity 99%
Jiang et al. [17]	—	Inception V3, Resnet152, Inception-Resnet-V2	Beijing Tongren Hospital	(Ensemble) Accuracy 88.21% Sensitivity 85.57% Specificity 90.85%
Tymchenko et al. [18]	—	EfficientNet-B4, EfficientNet-B5, SE- ResNeXt50	MESSIDOR, APTOS2019, EyePACS, IDRiD	(Ensemble) Accuracy 99% Sensitivity 99% Specificity 99%

detection. Then, only the output layer is replaced according to the given task and number of classes. Table I lists some of these studies along with the used pre-trained models and their performance. Whenever a study investigated more than one pre-trained model, an ensemble was applied to combine all these models and produce an optimal model.

Traditionally, the output layer of pre-trained models is replaced by a multi-layer neural network classifier and a softmax layer with a size equivalent to the number of classes to be recognised. Nevertheless, Taufiqurrahman et al.[11] suggested restructuring the MobileNetV2 model by replacing the fully connected layer with a Support Vector Machine (SVM) classifier. This modified version, MobileNetV2-SVM, obtained better performance than its original model. MobileNetV2-SVM achieved an accuracy of 85% and AUC of 92.8%. In a similar fashion, Khojasteh et al.[12] replaced the softmax layer with several classifiers: OPF, SVM, and KNN. Combing Resnet-50 and SVM outperformed other models with an accuracy of 98.2%, a sensitivity of 99%, and a specificity of 96%.

Several models can be combined using ensemble learning to improve prediction performance or reduce the bias in the learning process. Jiang et al. [17] introduced an image-based method to detect the DR early using an interpretable ensemble deep learning model. The proposed model is working on three main steps. Firstly, the fundus images preprocessing. Secondly, three different deep learning models have been used independently and trained sufficiently: Inception V3, Resnet152, and Inception-Resnet-V2. Finally, the Adaboost optimiser algorithm combined all the models' results to generate the final score. The integrated model proved a high performance in all evaluation metrics used: sensitivity, specificity, accuracy, AUC, 85.57%, 90.85%, 88.21%, and 0.946, respectively. Also, Tymchenko et al. [18] developed a DR detector using three-head CNN, which trained classification, regression and ordinal model. They used the output of these three heads for DR detection and achieved the sensitivity and specificity of 0.99.

As in the research, [17], [8], [6], [11], [7], this research aims to use InceptionV3, CNN, EfficientNetsB5 to detect DR due to their efficiency previous studies. However, these models will be validated using the same dataset to compare their results. Besides handling the issue of imbalanced distribution of classes on APTOS 2019 dataset, that not highlighted in previous researches.

III. Methodology

The purpose of this research is to classify a retinal fundus image whether it has a DR and at which severity. According to previous literature, deep learning and transfer learning models can solve this task. Transfer learning is a method that allows using the knowledge gained from other tasks to tackle new similar problems quickly and effectively. Hence, CNN models that are pre-trained will be fine-tuned utilizing domain dataset. Two pre-trained models will be selected for this study, InceptionV3 and EfficientNetsB5, for their effectiveness in diagnosing DR in the work of [17], [8], [6], [11], [7]. The performance of the fine-tuned pre-trained models will be compared with the performance of a CNN without pre-training.

A common issue in medical imaging datasets is the disparity in the number of samples within classes due to the difficulty of obtaining such samples. This problem is known as the class imbalance, and it pushes classifiers to prefer classes with higher training samples, reducing classification performance.

This section describes the techniques mentioned above.

A. Convolutional Neural Network (CNN)

Deep neural networks are artificial neural networks with more hidden layers to perform more complicated tasks and deal with massive amounts of data. The convolutional neural network (CNN) is one of the deep learning model networks with multiple layers such as convolution, pooling, fully connected, and non-linearity layer. CNN has been used with many applications, especially those that deal with spatial information, such as document analysis, image and video recognition, and computer vision [21]. The main aim of CNN is to increase or decrease the image dimensions into a more manageable form and extract the significant features, then process it to provide better predictions.

In this study, three convolution layers were employed with the same kernel size of (3,3). ReLU is used as an activation function with all layers, followed by a max-pooling layer with (2,2) pooling size to reduce the size of the large images. The results were flattened before the fully connected layer with a dropout of 0.2 to avoid overfitting. A softmax activation layer was used as the output layer. The architecture of the model is shown in Fig. 1. Some of the model's configurations were based on the work of [22], [23], [24].

B. Pre-Trained CNN Models

It has become customary to utilize a pre-trained CNN model and fine-tune it with in-domain dataset for the majority of computer vision applications. A pre-trained CNN model is a CNN model that has been trained on a large volume of data, such as ImageNet, for image classification [25]. Two pre-trained CNN models will be investigated in this paper: InceptionV3 and EfficientNetsB5.

Inception-v3 is a CNN architecture from the Inception family that contains 48 deep layers. Inception is characterised by implementing multiple kernels of different sizes in each layer (means become wider) instead of increasing the number of layers and going deeper in the network [26]. Each unit consists of four parallel operations: 1×1, 3×3, 5×5 conv layers and 3×3 max-pooling. All feature maps that come from different paths are concatenated together as the input of the next layer. Because in the image classification, the feature size of the image can diversify and deciding a fixed kernel size is difficult. Lager kernels are effective when the features are distributed over a wide area in the image. On the contrary, smaller kernels are useful and give excellent results in detecting small areas distributed across the image frame. To effectively recognise this variable size feature, kernels of different sizes are needed, which are provided in Inception models [27], [28].

EfficientNets family has a highly significant performance that achieves state-of-art on ImageNet, CIFAR-100, Flowers, and three other transfer learning datasets [29]. The architecture of the EfficientNets model involves convnet designs to reduce the space of the model with each layer to be scaled uniformly with a constant ratio to optimise the accuracy performance. It focuses on three aspects of scaling width, depth, and resolution. According to that, the EfficientNets family produces seven models with different image dimensions, and there is no change of layers operator of baseline network. This research proposes to apply EfficientNets B5 version.

C. Data Augmentation

Many approaches have been proposed to overcome the imbalanced dataset problem that can be classified into two categories: creating algorithms to resample the data and data preprocessing to generate new samples [30]. Resampling a dataset is a method used to balance the class distribution of the dataset. This is achieved by either adding samples to the minority class (oversampling) or removing samples from the majority class to balance the data (undersampling) [31]. However, data augmentation is a common technique used to generate new samples of the data to provide the image in a different representation.

Data augmentation techniques help improve the deep learning model's ability by generating artificial new images to achieve high variation in the training dataset and avoid overfitting problems. Many transform operations could be applied for data augmentation, such as random rotation, brightness, zoom, and image preprocessing techniques, such as Gaussian blur or CLAHE [32]. The data augmentation techniques included in this research are horizontal and vertical flip, zoom, and rotation. Fig. 2 shows a sample image from class 0 augmented after preprocessing image phase.

IV. Diabetic Retinopathy Dataset

Several datasets are available for the retina to detect DR and the vessels. Often these datasets are utilised for training,

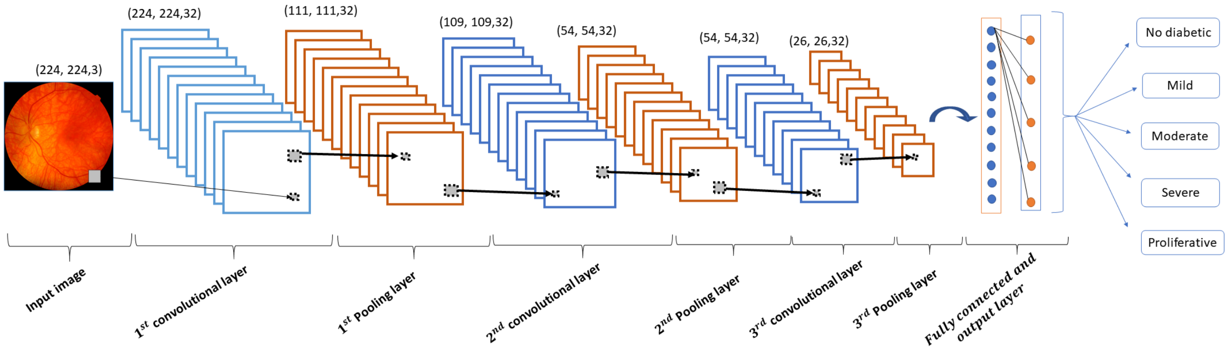


Fig. 1. CNN Module Architecture used in this Study.

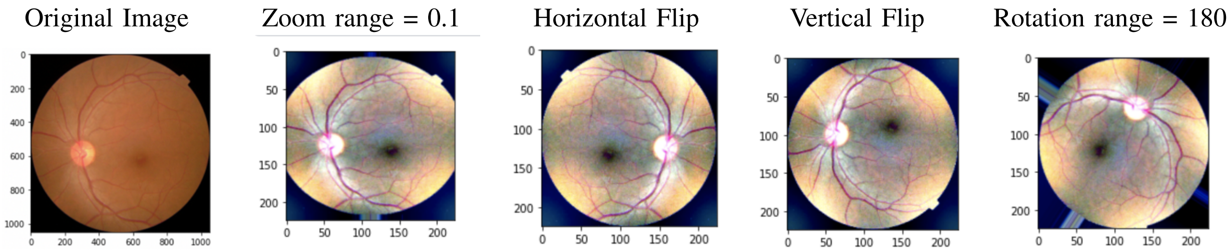


Fig. 2. Applying Data Augmentation Techniques in a Sample of Retinal Fundus Image.

validation, and testing deep learning models. The Asia Pacific Tele-Ophthalmology Society (APTOS) published this dataset in the second quarter of 2019. As shown in Table I, several studies used the APTOS2019 dataset [33] for blindness detection, containing a large set of retina images taken using fundus photography. Initially, two sets were published: labelled images, known as the train set, and unlabelled images, known as the test set. Only the labelled images were included in this study, consisting of 3662 fundus images. Each image was labelled into one of five classes, representing the severity of DR. Table II shows samples of each class and its characteristics that differentiate it from the others. As many of the medical dataset, APTOS2019 suffers from class imbalance as shown in Fig. 3 with majority of the cases towards healthy images without DR. However, there is a balance between the sum of all DR images regardless of their severity and healthy images.

The image size is a more critical factor that will impact the classification tasks. As shown in Fig. 4, there is a different distribution of image height and width, which suggests that not all images are in a perfect square shape.

V. Experiments

A. Experimental Design

All experiments were implemented and evaluated using Python [34] and leverage TensorFlow and Keras library [35] using Kaggle GPUs, Kaggle presents free access to NVidia K80 GPUs in kernels. In particular, these GPUs can be used to train deep learning models [33]. For this study, the labelled set was split into three homogeneous sets: training, validation and testing sets with a ratio of 68%, 20%, and 12%, respectively. The distribution of classes within each split is shown in Table III. Two sets of experiments were performed: fine-tuning and training using an imbalanced training set, 2489 samples, and a balanced training set after augmentation, 6158 samples.

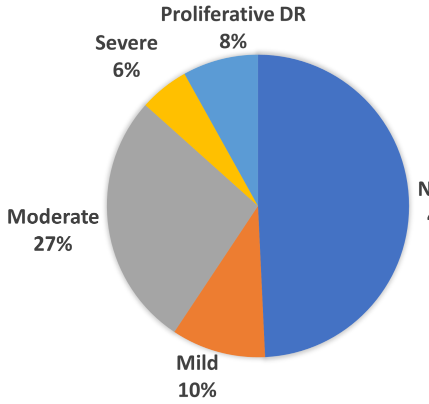
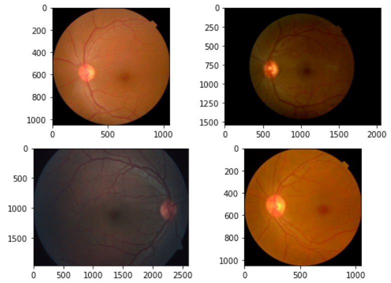
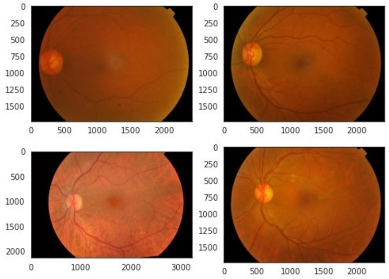
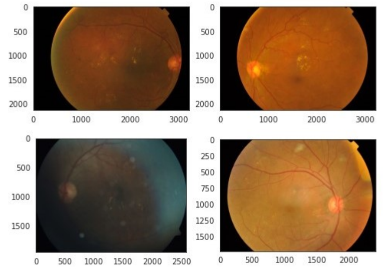
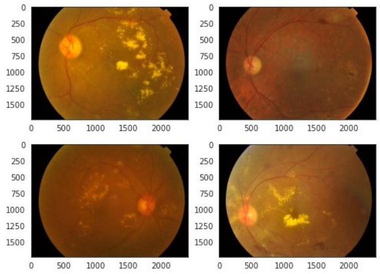
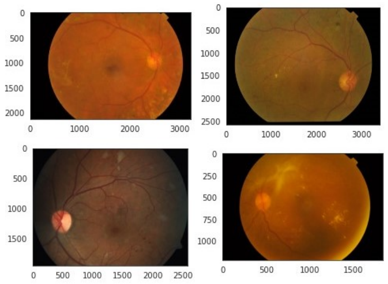


Fig. 3. Class Distribution among APTOS2019 Training Set.

B. Data Preprocessing

As most of the pre-trained models in this study were trained using images of size 224×224, images of APTOS2019 were rescaled accordingly. Moreover, images were converted into grayscale, which increases the visibility of some abnormalities. Following [18], [7], further image processing processes were applied: uninformative black areas were removed using circular crop, blending using Gaussian blur with alpha=4, beta=-4,

TABLE II. LABEL DESCRIPTION AND CHARACTERISTICS OF DR SEVERITY LEVELS WITH ITS SAMPLES FROM APTOS2019

Sample	Characteristics
 0 200 400 600 800 1000 0 500 1000 0 250 500 750 1000 1250 1500 0 500 1000 1500 2000 0 500 1000 1500 0 500 1000 1500 2000 2500 0 200 400 600 800 1000 0 500 1000	**Label 0: No diabetic retinopathy (NoDR)**
 0 250 500 750 1000 1250 1500 0 500 1000 1500 2000 0 250 500 750 1000 1250 1500 0 500 1000 1500 2000 0 500 1000 1500 2000 0 1000 2000 3000 0 250 500 750 1000 1250 1500 0 500 1000 1500 2000	**Label 1: Mild nonproliferative retinopath**: In this early stage of the disease, small patches of balloon-like swelling in the small blood cells in the retina, known as microaneurysms. The fluid will leak into the retinas through these microaneurysms as shown in the left images.
 0 500 1000 1500 2000 0 1000 2000 3000 0 500 1000 1500 2000 0 1000 2000 3000 0 500 1000 1500 0 500 1000 1500 2000 2500 0 250 500 750 1000 1250 1500 0 500 1000 1500 2000	**Label 2: Moderate nonproliferative retinopathy**: As the disease progresses, Blood vessels feeding the retina may swell and distort and also lose blood transportation capacity. These conditions cause significant changes to the appearance of the retina and can contribute to diabetic macular edema (DME) as shown in the left images.
 0 250 500 750 1000 1250 1500 0 500 1000 1500 2000 0 250 500 750 1000 1250 1500 0 500 1000 1500 2000 0 250 500 750 1000 1250 1500 0 500 1000 1500 2000 0 250 500 750 1000 1250 1500 0 500 1000 1500 2000	**Label 3: Severe nonproliferative retinopathy**: Many further blood vessels are blocked, which deprive the retinal region of blood supply. These regions secrete growth factors that suggest that the retina is forming new blood vessels as shown in the left images.
 0 500 1000 1500 2000 0 1000 2000 3000 0 500 1000 1500 2000 2500 0 1000 2000 3000 0 500 1000 1500 0 500 1000 1500 2000 2500 0 250 500 750 1000 0 500 1000 1500	**Label 4: Proliferative diabetic retinopathy (PDR):** This more serious form called proliferative diabetic retinopathy. Damaged blood vessels are blocked in the retina in this case, causing the development of irregular new blood vessels, and can flow into the clear, jelly-like substance that fills the center of the eye (vitreous). Scar tissue stimulated through new blood vessel growth can gradually separate the retina from the posterior of the eye. Therefore, retinal detachment could lead to permanent eyesight loss as shown in the left images.

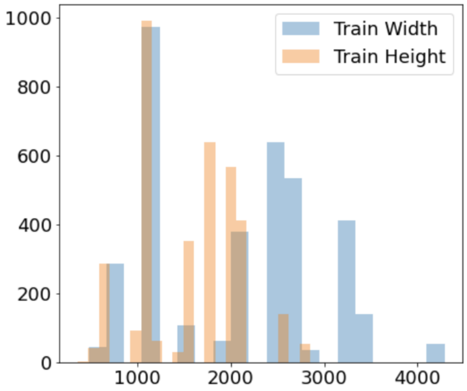


Fig. 4. Distribution of Image Width and Height in APTOS2019.

TABLE III. THE SPLIT OF APTOS2019 TRAIN SET USED IN THIS STUDY AND ITS CLASS DISTRIBUTION.

Class	Training set	Testing set	Validation set
No DR	1229	217	359
Mild	265	43	62
Moderate	670	124	205
Proliferative DR	132	25	36
Severe	193	31	71
Total	2489	440	733

and gamma=128. Consequently, the resulting images are not entirely greyscaled as modifications were applied separately on every pixel's colour channel. This helps improve the blood vessel's visibility and its growth in the eye, as shown in Fig. 5. All image preprocessing techniques was applied using Python (cv2) OpenCV library [36], [37].

C. Data Augmentation

Data augmentation was implemented using 'ImageDataGenerator' class from Keras library [35]. As shown in Fig. 3, the number of cases in each category varies significantly, with no_DR as the majority class (49.3% of total images). The number of the augmented images is different based on the number of the original images, as shown in Fig. 6. The augmentation phase enriched the diversities of the classes to provide high-quality images to the learning models. This operation was performed only on the training dataset. Image augmentation for the minority classes was applied via zooming, flipping, and rotation, which acquired a dataset three times larger than the original set.

D. Fine-Tuning the Pre-Trained CNN Models

For every pre-trained model included in this study, the input layer was set to be 224×224 and three channels. However, the output layer was modified to match the number of classes in this task, i.e. five classes. Then, all layers were frozen during the fine-tuning process except for the modified last layers. The last layers were trained using Adagrad optimiser with a learning rate of 0.01 and for 30 epochs. Similar training configurations were employed when training CNN model.

E. Evaluation Metrics

A multi-class classification task necessitates factors such as class balance and expected outcomes when picking the optimal metrics to evaluate the performance of a particular classifier against a given dataset. One performance metric may assess a classifier from a specific perspective while others can not, and vice versa. Hence, there is no standardised (unified) metric for defining the generalised performance measurement of the classifier. In this paper, several metrics are chosen to measure the models' performance: Accuracy, Precision, Recall and F1-score. Table IV. summarises how each of the first four metrics is calculated for a multi-class classifier with C classes, where TP_i and TN_i are the number of cases correctly diagnosed for class C_i or not, respectively. And FP_i and FN_i are the number of cases that were incorrectly diagnosed to the class C_i or not, respectively.

TABLE IV. EVALUATION METRICS AND THEIR FORMULAS.

Metric	Description	Formula
Accuracy	The average number of correct predictions.	$Acc = \frac{\sum_{i=1}^{C} \frac{TP_i+TN_i}{TP_i+FN_i+FP_i+TN_i}}{C}$
Precision	Capability of identifying the correct instances for each class.	$Pre = \frac{\frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C} TP_i+FP_i}}{C}$
Recall	Capability to recognise the true positive out of the total true positive cases.	$Rec = \frac{\frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C} TP_i+FN_i}}{C}$
F1-score	The harmonic average of precision and recall.	$F1 = 2 * \frac{Pre_M * Rec_M}{(Pre_M + Rec_M)}$

TABLE V. RESULTS WHEN TRAINING MODELS USING THE IMBALANCED DATASET.

	Precision	Recall	F1-score	Accuracy	kappa
	Training set				
CNN	61%	69%	69%	67%	61%
Inceptionv3	84%	71%	74 %	88%	84 %
EfficientNet B5	35%	34%	32%	62%	29%
	Validation set				
CNN	58%	65%	61%	65%	58 %
Inceptionv3	51%	51%	51%	76%	70%
EfficientNet B5	40%	36%	34%	64%	30%
	Testing set				
CNN	64%	71%	67%	73%	65%
Inceptionv3	62%	53%	54%	78%	72%
EfficientNet B5	30%	34%	31%	63%	30%

As one of the experiments uses an imbalanced dataset, Cohen's kappa was used as an additional metric. It can be computed as follows:

$$K = \frac{P_0 - P_e}{1 - P_e},$$

where P_0 denotes the overall accuracy and P_e denotes a measure of the probability of the agreement between the prediction class values and the actual class values as it occurs by chance [38]. $K = 1$ if classes are in complete agreement while $K = 0$ proves the opposite.

VI. RESULTS AND DISCUSSION

A. Training with Imbalanced Dataset

Each pre-trained model was fine-tuned using the imbalanced training set, with 2489 samples and no_DR as the majority class. When training the CNN model from scratch, the same imbalanced set was used for training. Table V. lists the results obtained during models training. Since accuracy is

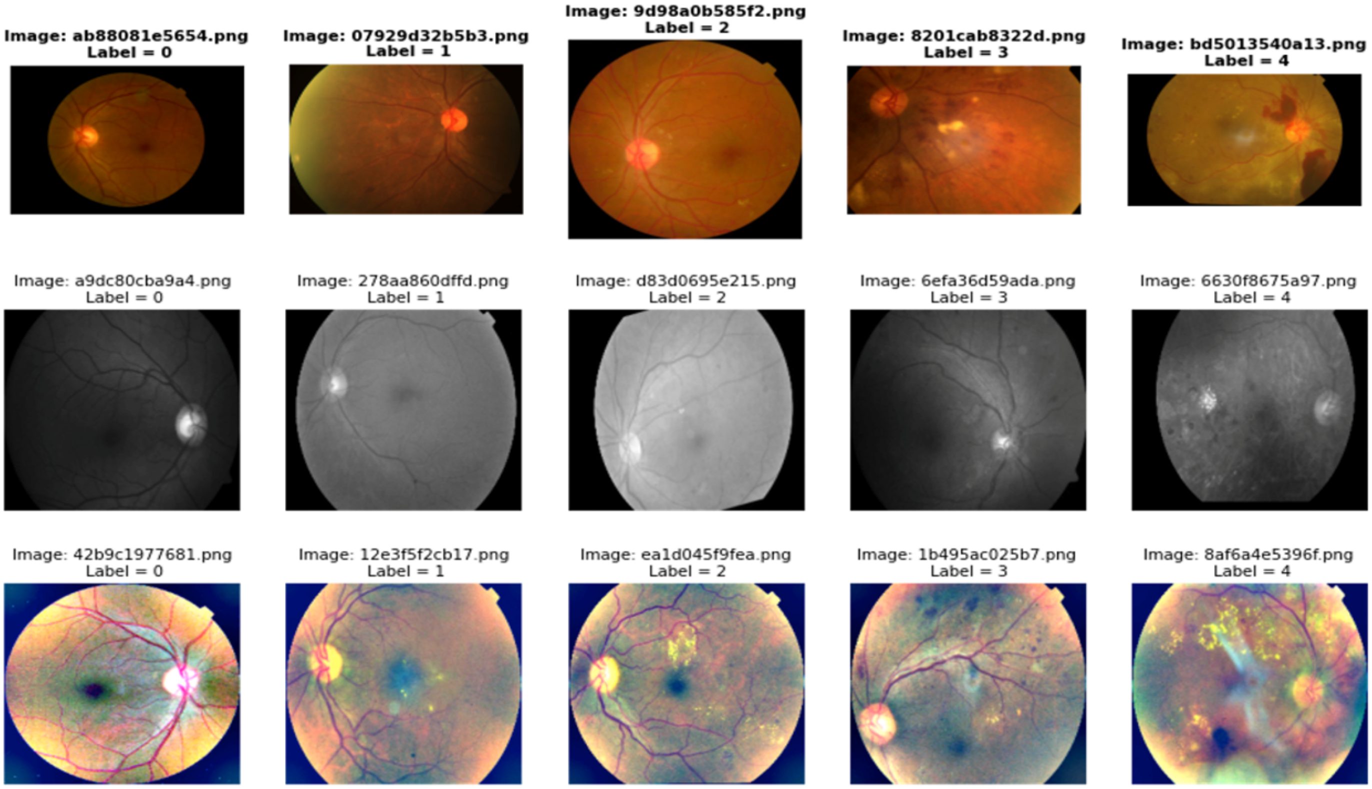


Fig. 5. Applying Image Preprocessing Techniques on Some Samples from APTOS2019. The First row Shows the Original Sample from each Class. The Second Row Shows the Same Samples after Converting them into Grayscale using the cvtColor Function and COLOR_BGR2GRAY as a Parameter. The Third Row Shows the Same Samples after Applying Gaussian Blur and Circular Cropping.

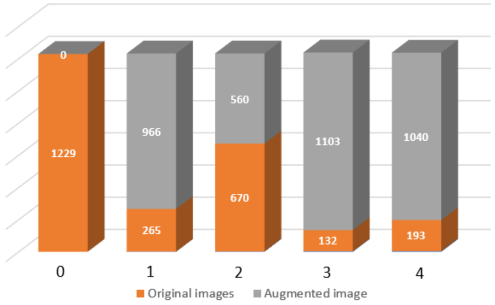


Fig. 6. Samples Per Class in APTOS2019 Training Set. Class 0 (no_DR) is the Majority Class so Other Classes were Augmented with Different Numbers of Samples to Obtain a Balanced Training Set.

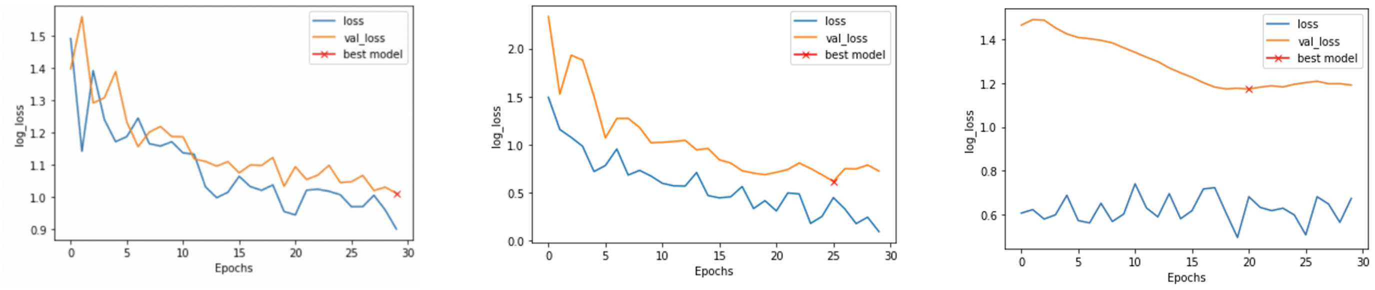
unreliable when evaluating models trained on an imbalanced dataset, F1-score and kappa are the primary evaluating metric. CNN model achieves the highest F1-score with 67%, while the InceptionV3 model obtained 54%. On the other hand, the EfficientNetB5 model has the lowest performance.

To investigate the reasons for EfficientNetB5 performance, the learning curve for each of these models are depicted in Fig. 7. As shown in the figures, the learning curves of the CNN and InceptionV3 model in training and validation phases was improving smoothly, while the EfficientNetB5 model suffered from a high overfitting problem, which caused its low results.

Fig. 8 visualises the confusion matrix of these models, which indicates the number of predictions produced by the model where it classified the classes correctly or incorrectly. The diagonal expresses the correctly diagnosed states for each class, where the off-diagonal elements represent the misclassified samples. In general, all models have their best recognition for Class 0 (no_DR) and 2 (mild NPDR) aligned with the class majority shown in Fig. 3. with Class 0 and 2 with the largest samples, respectively. However, most confusion was between different classes of DR, not no_DR and any DR. This observation was accurate for all models. In other words, these models have good DR detection but poor severity level classification. The detection rate can be calculated by mapping all DR severity levels 1-4 to 1. Hence, the obtained detection rates are 90%, 96% and 83% for CNN, InceptionV3 and EfficientNetB5, respectively.

B. Training with Balanced Dataset

In this experiment, pre-trained models were fine-tuned using the balanced training set via augmentation, with 6158 samples. The same set was used for training when training the CNN model from scratch. Table VI lists the results obtained during models training. CNN model achieves the highest F1-score with 64%, while the InceptionV3 model obtained 58%. On the other hand, the EfficientNetB5 model has the lowest performance with a 48% F1-score. Looking at the learning curves for these models in Fig. 9. the performance of the validation set improved better than the training set for the CNN

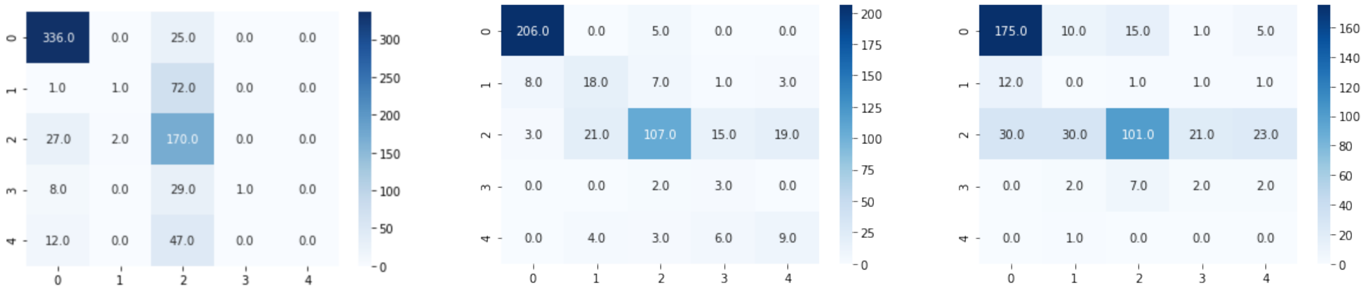


(a) CNN

(b) InceptionV3

(c) EfficientNetB5

Fig. 7. Learning Curves for the Models Trained using the Imbalanced Dataset using the Train Set (2489 Samples) and the Validation Set (733 Samples).



(a) CNN

(b) InceptionV3

(c) EfficientNetB5

Fig. 8. The Confusion Matrix when Evaluating on the Test Set (440 Samples) for the Models Trained using the Imbalanced Dataset. The x-Axis Represents the Actual Labels, while the y-axis Represents Predicted Labels. Label 0: no_DR, Label 1: mild NPDR, Label 2: Moderate NPDR, Label 3: Severe NPDR, and Label 4: PDR.

TABLE VI. RESULTS WHEN TRAINING MODELS USING THE BALANCED DATASET

	Precision	**Recall**	**F1-score**	**Accuracy**	**kappa**
			Training set		
CNN	50%	50%	49%	50%	28%
Inceptionv3	76%	75%	75%	75%	64%
EfficientNet B5	59%	59%	59%	59%	33%
			Validation set		
CNN	57%	65%	61%	65%	53%
Inceptionv3	64 %	60%	60%	79%	69%
EfficientNet B5	59%	47%	47%	74%	62%
			Testing set		
CNN	61%	68%	64%	68%	57%
Inceptionv3	59%	59%	58%	78%	70%
EfficientNet B5	57%	46%	48%	73%	60%

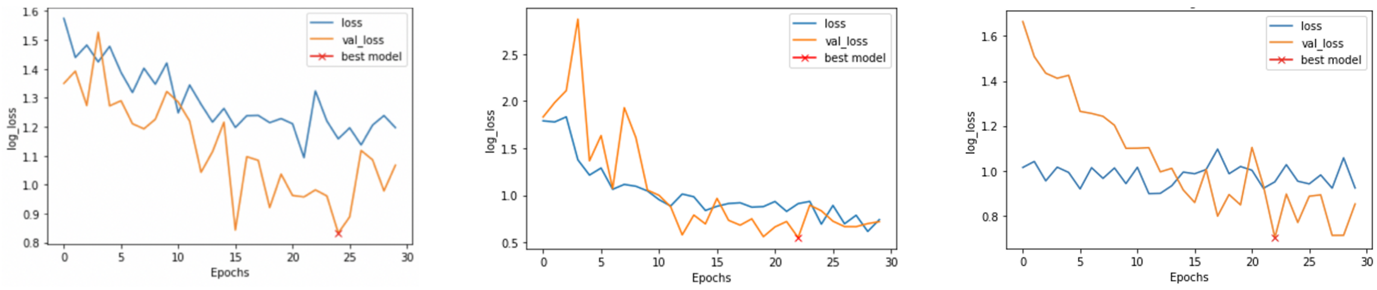
model, which indicates that some samples were difficult for the models to learn from the features. However, this was not observed for InceptionV3 and EfficientNetB5 models, which means the performance of training and validation sets were approximate are similar.

Fig. 10 visualises the confusion matrix of these models. In general, all models could not recognise Class 4 (PDR) successfully compared to other classes. As in the previous experiment, most confusion was between different classes of DR, not no_DR and any DR. The obtained detection rates here are 84%, 95% and 90% for CNN, InceptionV3 and EfficientNetB5, respectively.

This study performed two experiments; the first was on a dataset imbalanced between classes and only processed by scaling and resizing the image. The second experiment was on a balanced dataset by utilising augmentation data and applying image preprocessing techniques. F1-score was used to measure and compare the performance in both experiments because it is a standard measure of imbalanced data classification, in addition to the rest metrics mentioned in Section V-E. The performance was improved when using balanced since the InceptionV3 and EfficientNetB5 models obtained higher results. InceptionV3 model's performance improved in Recall and F1-score when using a balanced training set while the results of the CNN model decreased in all measures. On the other hand, the results of the EfficientNetB5 model improved in all metrics when using a balanced training set. Hence, fine-tuning pre-trained models could benefit from the augmented samples and enhanced features, which was not the case for the CNN model.

Furthermore, the CNN model achieves the highest results in the two experiments which are 67% and 64% of F1-score, in the first and second experiments, respectively. When looking at their learning curves, overfitting was an issue in pre-trained models, indicating the need for more powerful regularisation for these advanced architectures. In other words, the more complex the architecture, the more prone to overfit.

In general, the detection ability of these models was better than its classification between DR severity levels. For EfficientNetB5, the DR detection was improved by 7% absolute when

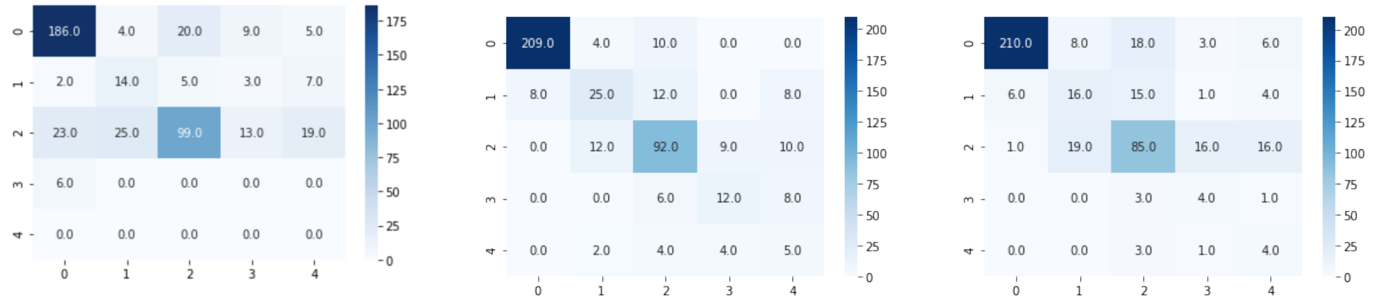


(a) CNN

(b) InceptionV3

(c) EfficientNetB5

Fig. 9. Learning Curves for the Models Trained using the Balanced Dataset using the Augmented Train Set (6158 Samples) and the Validation Set (733 Samples).



(a) CNN

(b) InceptionV3

(c) EfficientNetB5

Fig. 10. The Confusion Matrix when Evaluating on the Test Set (440 Samples) for the Models Trained using the Augmented Balanced Dataset. The x-Axis Represents the Actual Labels, while the y-Axis Represents Predicted Labels. Label 0: no_DR, Label 1: mild NPDR, Label 2: Moderate NPDR, Label 3: Severe NPDR, and Label 4: PDR.

using a balanced training set, while it was the opposite case for CNN as its detection accuracy dropped by 6% absolute.

VII. Conclusion

DR is currently one of the dominant diseases that significantly affect people with diabetes. The paper covers the details of the implementation and evaluation of several deep learning models: CNN, InceptionV3, and EfficientNetsB5 for classifying DR using the APTOS2019 dataset. Two different experiments were conducted, the first with the original images and the second after processing the images and balancing the classes. The InceptionV3 model performed the best accuracy on the dataset in both experiments, while the CNN model got the highest F1-score in both experiments. Using these prediction results, effective DR detection systems can be implemented using deep learning models so that the patient can be treated and dealt with in the early stages. The results of this research may not be the same as previous research due to the difference in the dataset used and the data processing method. This research's main challenges and limitations are that the image dataset was imbalanced, and there was a shortage of efficiency of the devices utilised in processing even when using online GPU, such as Kaggle, the allotted time was limited. For further work, this research can expand to address these deficiencies by using other methods to balance data and apply other pre-trained models to diagnose DR.

Acknowledgment

The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (22UQU4260137DSR02)

References

[1] Mayo Clinics. Diabetic retinopathy. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/symptoms-causes/syc-20371611, [Accessed: May 25, 2021]

[2] L. Chen, D. J. Magliano, and P. Z. Zimmet, "The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives," *Nature reviews endocrinology*, vol. 8, no. 4, p. 228, 2012.

[3] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[4] NHS. Overview-Diabetic retinopathy. [Online]. Available: https://www.nhs.uk/conditions/diabetic-retinopathy/, [Accessed: 2022-Jan-01]

[5] Q. H. Nguyen, R. Muthuraman, L. Singh, G. Sen, A. C. Tran, B. P. Nguyen, and M. Chua, "Diabetic retinopathy detection using deep learning," in *Proceedings of the 4th International Conference on Machine Learning and Soft Computing*, 2020, pp. 103–107.

[6] S. Masood, T. Luthra, H. Sundriyal, and M. Ahmed, "Identification of diabetic retinopathy in eye images using transfer learning," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*. IEEE, 2017, pp. 1183–1187.

[7] M. M. S. Maswood, T. Hussain, M. B. Khan, M. T. Islam, and A. G. Alharbi, "Cnn based detection of the severity of diabetic retinopathy from the fundus photography using efficientnet-b5," in *2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2020, pp. 0147–0150.

[8] E. AbdelMaksoud, S. Barakat, and M. Elmogy, "Diabetic retinopathy grading system based on transfer learning," *arXiv preprint arXiv:2012.12515*, 2020.

[9] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. A. Khan, and W. Jadoon, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150 530–150 539, 2019.

[10] J. Gao, C. Leung, and C. Miao, "Diabetic Retinopathy Classification Using an Efficient Convolutional Neural Network," *Proceedings - 2019 IEEE International Conference on Agents, ICA 2019*, pp. 80–85, 2019.

[11] S. Taufiqurrahman, A. Handayani, B. R. Hermanto, and T. L. E. R. Mengko, "Diabetic retinopathy classification using a hybrid and efficient mobilenetv2-svm model," in *2020 IEEE REGION 10 CONFERENCE (TENCON)*, 2020, pp. 235–240.

[12] P. Khojasteh, L. A. Passos Júnior, T. Carvalho, E. Rezende, B. Aliahmad, J. P. Papa, and D. K. Kumar, "Exudate detection in fundus images using deeply-learnable features," *Computers in Biology and Medicine*, vol. 104, no. July 2018, pp. 62–69, 2019. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2018.10.031

[13] D. J. Hemanth, O. Deperlioglu, and U. Kose, "An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network," *Neural Computing and Applications*, vol. 32, no. 3, pp. 707–721, 2020. [Online]. Available: https://doi.org/10.1007/s00521-018-03974-0

[14] U. Kose, O. Deperlioglu, J. Alzubi, and B. Patrut, "Diagnosing diabetic retinopathy by using a blood vessel extraction technique and a convolutional neural network," *Studies in Computational Intelligence*, vol. 909, pp. 53–72, 2021.

[15] H. N. Pham, R. J. Tan, Y. T. Cai, S. Mustafa, N. C. Yeo, H. J. Lim, T. T. Do, B. P. Nguyen, and M. C. H. Chua, "Automated grading in diabetic retinopathy using image processing and modified efficientnet," in *International Conference on Computational Collective Intelligence*. Springer, 2020, pp. 505–515.

[16] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognition Letters*, vol. 133, pp. 210–216, 2020. [Online]. Available: https://doi.org/10.1016/j.patrec.2020.02.026

[17] H. Jiang, K. Yang, M. Gao, D. Zhang, H. Ma, and W. Qian, "An interpretable ensemble deep learning model for diabetic retinopathy disease classification," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 2045–2048.

[18] B. Tymchenko, P. Marchenko, and D. Spodarets, "Deep learning approach to diabetic retinopathy detection," *arXiv preprint arXiv:2003.02261*, 2020.

[19] J. Shan and L. Li, "A deep learning method for microaneurysm detection in fundus images," in *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2016, pp. 357–358.

[20] A. Singh and W. Kim, "Detection of diabetic blindness with deep-learning," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020, pp. 2440–2447.

[21] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*. Ieee, 2017, pp. 1–6.

[22] hman. How to add a reshape layer to the start of a pre-trained cnn. [Online]. Available: https://stackoverflow.com/questions/61742075/how-to-add-a-reshape-layer-to-the-start-of-a-pre-trained-cnn, [Accessed Feb 2021]

[23] P. Huilgol. Top 4 pre-trained models for image classification with python code. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/08/top-4-pre-trained-models-for-image-classification-with-python-code/, [Accessed Feb 2021]

[24] A. H. Inception model with custom input tensor. [Online]. Available: https://groups.google.com/g/keras-users/c/G9C55N7e8S4?pli=1/, [Accessed March 5, 2021]

[25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[27] A. Anwar. Difference between AlexNet, VGGNet, ResNet, and Inception. [Online]. Available: https://towardsdatascience.com/the-w3h-of-alexnet-vggnet-resnet-and-inception-7baaaecccc96, [Accessed: 2022-Jan-01]

[28] S.-H. Tsang. Review: GoogLeNet (Inception v1) Winner of ILSVRC 2014 Image Classification. [Online]. Available: https://medium.com/coinmonks/paper-review-of-googlenet-inception-v1-winnr-of-ilsvlc-2014-image-classification-c2b3565a64e7, [Accessed: 2022-Jan-01]

[29] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th International Conference on Machine Learning, ICML 2019*, vol. 2019-June, pp. 10 691–10 700, 2019.

[30] D. Ramyachitra and P. Manikandan, "Imbalanced dataset classification and solutions: a review," *International Journal of Computing and Business Research (IJCBR)*, vol. 5, no. 4, pp. 1–29, 2014.

[31] V. Ganganwar, "An overview of classification algorithms for imbalanced datasets," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 4, pp. 42–47, 2012.

[32] J. Brownlee, "How to Configure Image Data Augmentation in Keras." [Online]. Available: https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/, [Accessed: 2022-Jan-01]

[33] A. P. T.-O. Society. APTOS 2019 Blindness Detection dataset. [Online]. Available: https://www.kaggle.com/c/aptos2019-blindness-detection, [Accessed: 2022 Jan, 01]

[34] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[35] F. Chollet *et al.* Keras. [Online]. Available: https://github.com/fchollet/keras, [Accessed: 2022 Jan, 01]

[36] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[37] Divakar. Crop black border of image using numpy. [Online]. Available: https://codereview.stackexchange.com/a/132934, [Accessed: Feb 25, 2021]

[38] M. Widmann. Cohen's Kappa. [Online]. Available: https://www.knime.com/blog/cohens-kappa-an-overview, [Accessed: 2022-Jan-01]

Real-time Egyptian License Plate Detection and Recognition using YOLO

Ahmed Ramadan Youssef
Computer Science
Fayoum University
Egypt

Abdelmgeid Ameen Ali
Computer Science
Minia University
Egypt

Fawzya Ramadan Sayed
Computer Science
Fayoum University
Egypt

***Abstract*—Automatic License Plate Detection and Recognition (ALPR) is one of the most significant technologies in intelligent transportation and surveillance across the world. It has many challenges because it affects by many parameters such as the country's layout, colors, language, fonts, and several environmental conditions so, there isn't a consolidated ALPR system for all countries. Many ALPR methods have been proposed based on traditional image processing and machine learning algorithms since there aren't enough datasets, particularly in the Arabic language. In this paper, we proposed a real-time ALPR system for the Egyptian license plate (LP) detection and recognition using Tiny-YOLOV3. It consists of two deep convolutional neural networks. The experimental results in the first available publicly Egyptian Automatic License Plate (EALPR) dataset show the proposed system is more robust in detecting and recognizing the Egyptian license plates and gives mean average precision values of 97.89% and 92.46% for LP detection and character recognition, respectively.**

***Keywords*—Automatic license plate recognition; Egyptian license plate; Tiny-YOLOV3; CNN; eALPR dataset**

I. Introduction

Numerous applications, including traffic surveillance, automatic toll collection, parking, and theft prevention, depend on the automatic detection and recognition of licence plates [1]. The main objective of ALPR systems is to detect the location of the license plate in the vehicle image and recognize its characters and digits without involving massive human resources, and saving processing time.

The ALPR problem has been investigated over the last decades by many researchers, and several methods have been presented. However, the ALPR is still active due to several difficulties that are related to digital image processing, including the diversity of lighting, camera angles, distance from the camera, scales, complex backgrounds, etc. In addition to the license plate layout, each country defines a special plate structure, background, color, language, font type, and size, which differ from country to country. The majority of ALPR proposed methods are outside the Middle East (e.g. United States, China, Brazil, Taiwan, and Europe) [2]–[5].

Most of the previous works are based on traditional image processing techniques such as binarization, edge-based, character-based, color-based, texture-based, and template matching approaches, which are based on hand-engineered features. With the rapid development of the deep learning, the prediction performance of many computer vision tasks such as image classification, segmentation, and object detection has been greatly improved [6]. Many approaches have used deep Convolutional Neural Networks (CNN) to obtain state-of-the-art accuracy. However, the main drawback of deep learning models is that they need big datasets for the training stage. so, few deep learning works were presented in the Arabic region due to the lack of a publicly available dataset, especially for Egyptian license plates.

You Only Look Once (YOLO) [7] inspired models have made significant advancements in object detection problems. So we decided to use it in our proposed system. YOLOV3 [8] is the enhanced of YOLO and YOLOV2 [9] real time object detection algorithm that uses a model with 53 fully convolutional layers on the other side, Tiny-YOLOV3 is a light model focused on a speed / accuracy trade-off that uses eight convolutional and six max-pooling layers. Hence, Tiny-YOLOV3 is quicker but less accurate than YOLOv3.

In this paper, we propose a new real-time system for Egyptian license plate detection and recognition based on the Tiny-YOLOv3 object detection deep learning model. It consists of two stages : the first stage is responsible for detecting the license plate location from the input vehicle image and the second stage is recognizing the characters and digits from the detected license plate. The proposed system is used to evaluate the EALPR dataset that the first public Egyptian LP dataset.

our main contribution can be summarized as follows:

- A new real-time system for Egyptian license plate detection and recognition based on the Tiny-YOLOv3 deep learning model for both stages.
- Evaluation of the proposed system on the EALPR dataset, including 2450 vehicle images and more than 12,000 Arabic digits and characters.

The paper is organized as follows. We briefly review ALPR related works in Section II. Section III introduces the EALPR proposed system. We report the experimental results in Section IV, and finally, we present the conclusions and future work in Section V.

II. Related Work

ALPR is challenging because it depends on several factors, including the plate country's design, colors, environment, language, and texture. Both license plate detection and character recognition could be done using the object detection techniques. There are two approaches of object detection features, hand-engineered features (Haar, HOG, SIFT, SURF, ... etc.) and deep neural networks features based on convolutional neural networks (YOLO, SSD, Faster-RCNN, RefineDet, ... etc.). We briefly review some previous related work to the ALPR problem.

In [10], proposed a smart vehicle control system based on ALPR. The proposed method is color-based and extracts the region of the license plate based on some of the predefined colors like yellow and white, and applies histogram algorithms and morphological operations for rapid detection. This method uses a back-propagation neural network for character recognition. A total of 350 vehicles are used for testing the method and achieving a high success rate.

In [11], presented an ALPR system for the Egyptian license plates. A total of 221 images of classic Egyptian vehicle license plates were used in this study. The proposed method accuracy for plate detection is 78% using edge detection and morphological operations, 85% using the illumination model of YCbCr and wavelet transform, 85% for character segmentation, and 74% for character recognition.

In [2], developed an end-to-end deep learning system for Brazilian license plate detection and recognition in unconstrained scenarios. The developed system is based on deep convolutional neural networks (CNN). The authors used the public available Brazilian datasets. The accuracy of the license plate detection is 99%, and the recognition is 93%.

In [5], presented a two stages method for license plate detection and recognition. In the first stage, the author detected the vehicle region using a faster R-CNN method and in the second stage, they used the same algorithm with hierarchy sampling method for plate detection in the vehicle region. The proposed method is evaluated by Caltech dataset and achieved 98.39% as the precision rate.

In [3], proposed a real-time end-to-end ALP system based on the state-of-the-art YOLO object detection. They presented a fully annotated public dataset called FPR-ALP for Brazilian license plates. Fast-YOLO, YOLOV2, and CR-Net are used for plate detection and recognition respectively. The recognition accuracy of the system is 93.53% higher than both the commercial Sighthound and OpenALPR solutions.

III. Proposed System

A. System Overview

In Fig. 4, we show the main stages of the proposed ALPR system. The development of real-time object detection techniques is an important step for building a reliable ALPR system. We use two object detection models for license plate detection and recognition. The input to our proposed system is the vehicle image, whereas the output is the recognized license plate characters and digits. Tiny-YOLOV3 is the backbone architecture for both stages LP detection and recognition.

B. ELAPR Dataset

In this section, we introduce a new dataset for the Egyptian license plate called (EALPR). We scrap the EALPR dataset images from websites such as Instagram pages [12]–[15] and Facebook Marketplace [16] using web scraping Tools [17], [18]. It has 2,450, and 12,160 characters images including many types of vehicles such as cars, buses, trucks, microbuses, and mini-busses. Fig. 1 shows the characters and digits statistics. The vehicle images are captured using different cameras, at varying times, lighting, and background. Fig. 7 shows samples from EALPR. It is publicly available on https://github.com/ahmedramadan96/EALPR.

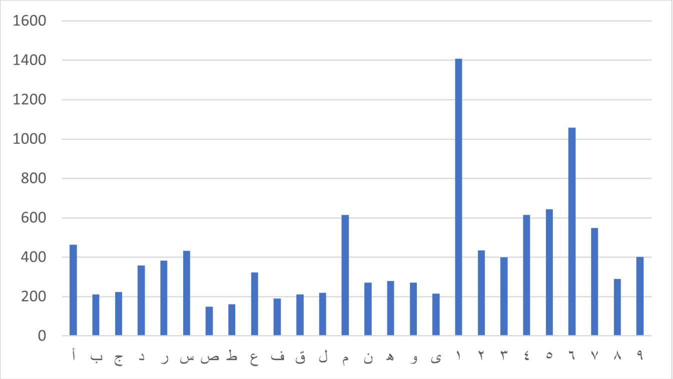


Fig. 1. Characters and Digits Statistics in the EALPR Dataset.

The Egyptian License Plate (LP) uses the Arabic digits and characters which is varying in several countries that use Latin letters such as Europe, Brazil, US, ... etc. It has a dimension 16 (height) X 40 (width) with aspect ratio 1:2 approximately [19]. its layout is divided into 3 regions: the country region contains the Egypt word in Arabic and English format, the character region includes plate characters, and the number region contains plate digits. Fig. 2 shows the Egyptian LP structure. It uses 10 Arabic digits and 17 Arabic characters. Fig. 3 illustrates the used license plate Arabic characters, digits, and their corresponding English format.

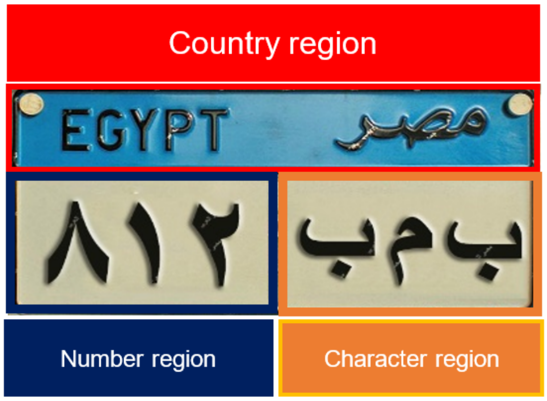
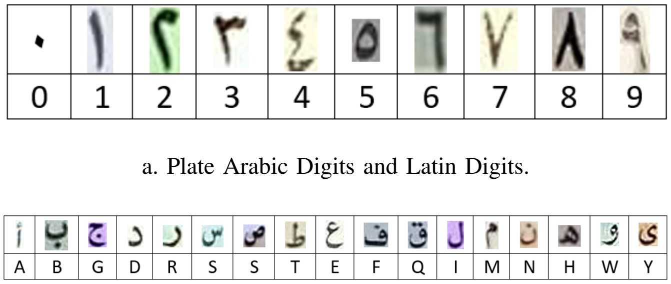


Fig. 2. Egyptian License Plate Structure.

Dataset annotation is required for object detection YOLO models. After collecting EALPR dataset we manually annotate the vehicles, characters, and digits using Ybat tool [20] as shown in Fig. 5, and 6. Ybat provide the start and end coordinates (x_{min},x_{max}), (y_{min},y_{max}) respectively for each rectangle. The four values of YOLO bounding boxes is calculated using



a. Plate Arabic Digits and Latin Digits.

b. Plate Arabic Characters and Latin Characters.

Fig. 3. a. Plate Digits b. Plate Characters in Arabic (First Row) & English (Second Row).

equations (1)-(4)

$$b_{\mathrm{x}} = \frac{x_{\min} + x_{\max}}{2 \times w} \tag{1}$$

$$b_{\mathrm{y}} = \frac{y_{\min} + y_{\max}}{2 \times h} \tag{2}$$

$$b_{\mathrm{w}} = \frac{x_{\max} - x_{\min}}{w} \tag{3}$$

$$b_{\mathrm{h}} = \frac{y_{\max} - y_{\min}}{h} \tag{4}$$

where b_x,b_y,b_w,b_h representing the center point of the rectangle, width, and height, respectively.

C. *License Plate Detection Network **(LP-Net)***

LP-Net is responsible for taking the raw vehicle image as input and detecting the location of the plates. The plates are cropped based on the bounding boxes using OpenCV [21] and passed to the character model. A license plate is an object that may be detected using object detection algorithms. The main function of object detection algorithms is to detect the object locations at different scales (shapes and sizes). Deep learning object detection techniques work effectively in a variety of environments and with a large dataset. LP-Net is the same original Tiny-YOLOV3 network. Tiny-YOLOV3 consists of 8 convolutional and 6 max-pooling layers in the feature extractor block. The network architecture is shown in Table I. It uses small kernel sizes 3×3 and 1×1 for the convolutional layers and 2×2 for max-pooling layers. We changed the input image size from 416×416 to 608×608, which resulted in higher accuracy. We also changed the final convolutional layer to predict only one class label ("license_plate"). YOLO predicts bounding boxes using A anchor boxes (we choose A = 3) with four values (b_x,b_y,b_w,b_h), confidence, and C class probability, therefore the number of filters is 18 which is defined in Equation (5)

$$\#Filters = (Classes + 5) \times Anchors \tag{5}$$

TABLE I. Tiny-YOLOV3 Architecture

Layer	Type	Filters	Size/stride
0	Conv	16	3×3/1
1	Max		3×3/2
2	Conv	32	3×3/1
3	Max		3×3/2
4	Conv	64	3×3/1
5	Max		3×3/2
6	Conv	128	3×3/1
7	Max		3×3/2
8	Conv	256	3×3/1
9	Max		3×3/2
10	Conv	512	3×3/1
11	Max		3×3/2
12	Conv	1024	3×3/1
13	Conv	256	3×3/1
14	Conv	512	3×3/1
15	Conv	33	3×3/1
16	Yolo loss		
17	Route 13		
18	Conv	128	3×3/1
19	Upsampling		×2
20	Route 19,8		
21	Conv	256	3×3/1
22	Conv	33	3×3/1
23	Yolo loss		

D. *Character Recognition Network **(Char-Net)***

The main function of Char-Net is to detect the plate digits and characters and recognize them using deep object detection CNN. The input for this network is the cropped license plates that were detected from the previous network. We accept all cropped (LP)s with an aspect ratio of 1:2 and reject the small plates. Also, It has the same architecture of Tiny-YOLOV3 as shown in Table I with some changes. All plates are resized to be 384×192 to prevent the vanishing gradient in CNN layers. The architecture is changed to detect 27 digits and characters that are used in the Egyptian license plates. Fig. 3 show all used digits and character classes. The number of filter in the last convolutional layer is 96 which calculated using Equation 5

IV. Experimental Results and Evaluation

A. *Experimental Setup*

All experiments were performed on a personal computer with Windows 10 64-bit, Intel(R) Core(TM) i7-9750H 2.6GHZ CPU, 16GB memory, and NVIDIA GeForce GTX 1660Ti GPU. The proposed system is implemented using the darknet deep learning framework and OpenCV library. The YOLO implementation is available here [22]. As previously mentioned, the EALPR dataset vehicle images are collected from social networks with different environments and conditions. For plate and character networks, we resize the input image to be 608×608 and 384×192 respectively. 80% of the dataset is selected randomly for training our proposed system and the remaining 20% is used for testing purposes. In the training stage, we use the pre-trained Tiny-YOLOV3 model to initialize the network's weights. Both the plate and character networks were trained for 5000 epochs and 6000,54000 max batches respectively with a batch size of 64 images. We use the Leaky RELU activation function. The momentum is set to 0.9, the weight decay to 0.0005, and the learning rate started with 0.001 used for both networks. Table II summarizes all training parameters for both models.

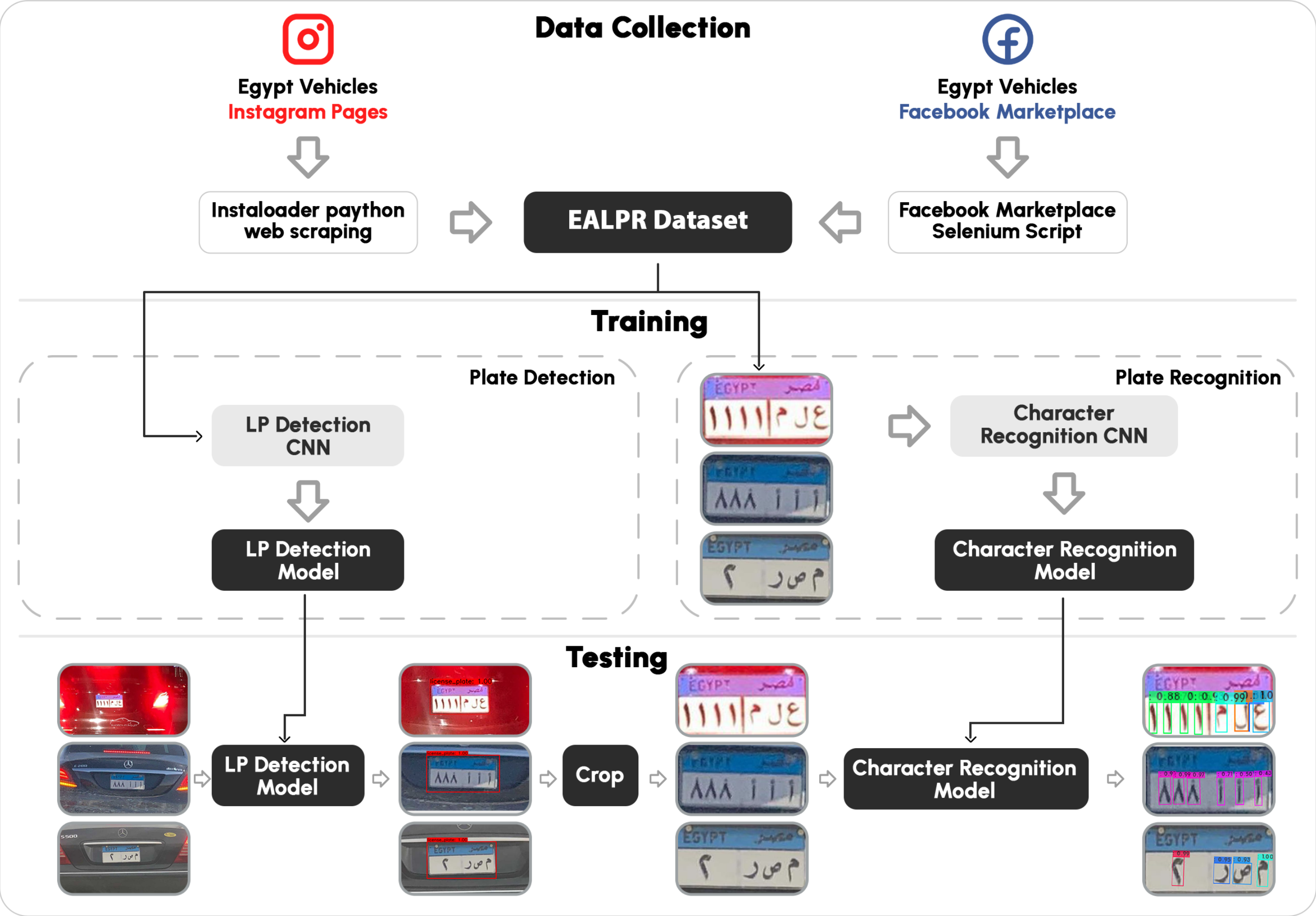


Fig. 4. Main Stages of the Proposed EALPR Approach.

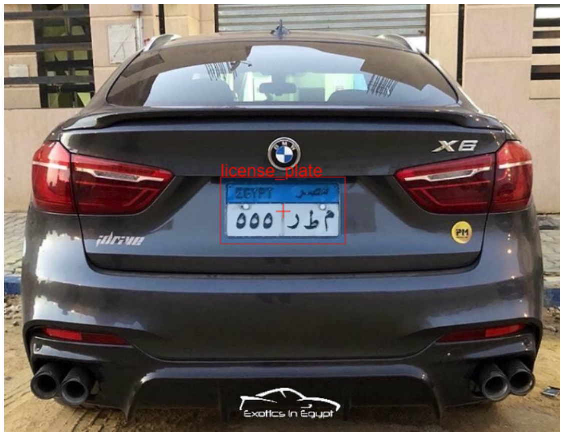


Fig. 5. Vehicle Plate Annotation.

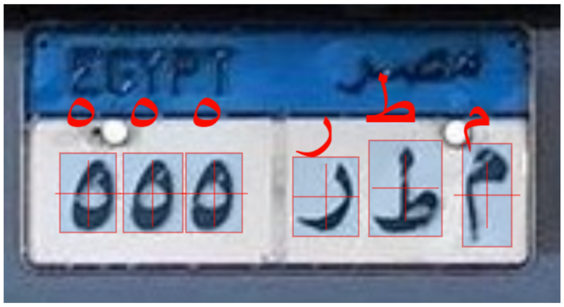


Fig. 6. Plate Characters Annotation.

B. Evaluation Metrics

The proposed system is evaluated using the mean average precision (mAP) value. The mAP value depends on precision and recall metrics [23] which are described in Equations (6) and (7).

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

where True Positive(TP), False Positive(FP), and False Negative(FN) means the number of correctly detected (plates

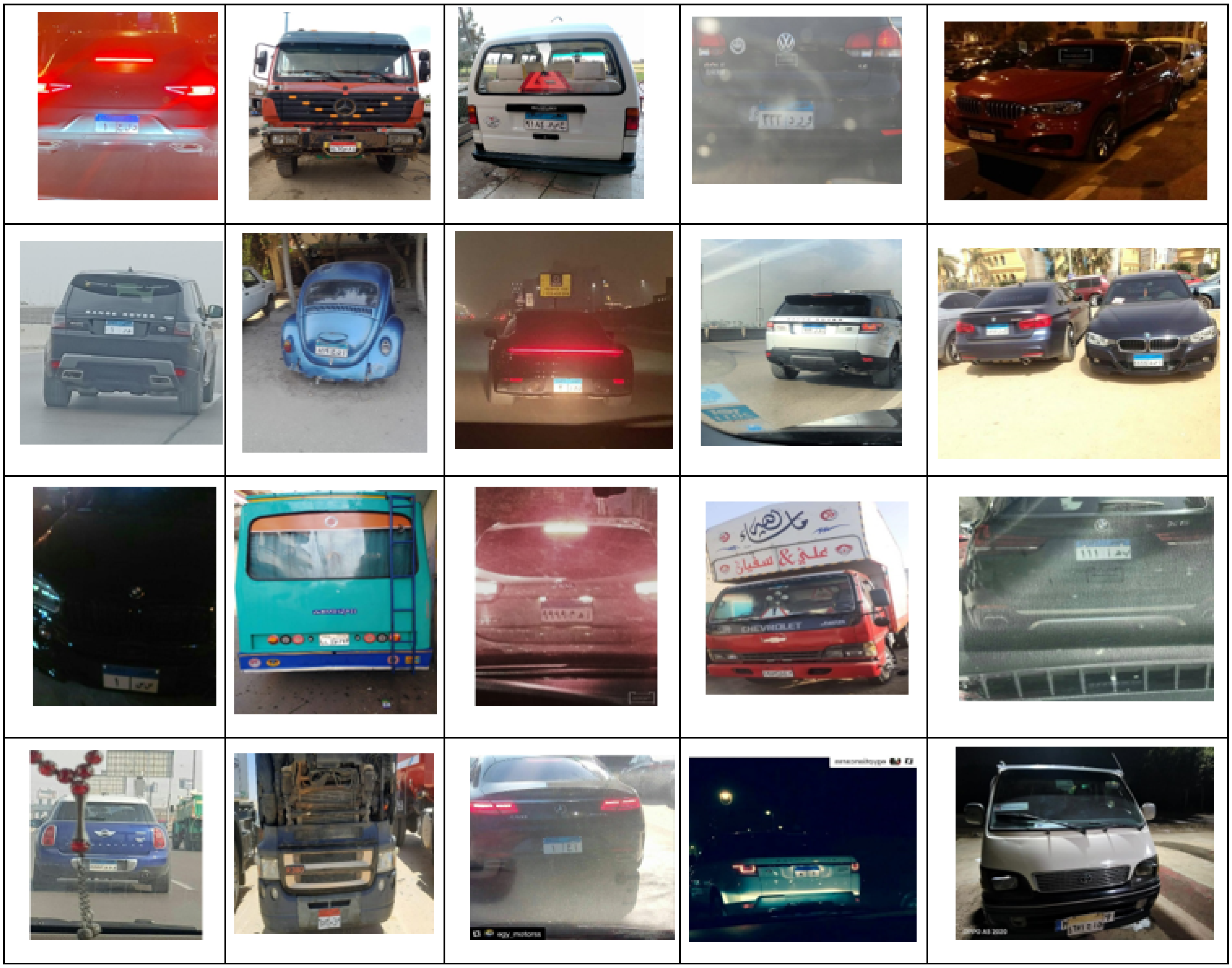


Fig. 7. Sample Images from EALPR Dataset.

TABLE II. TINY-YOLOV3 TRAINING PARAMETERS FOR LP-NET & CHAR-NET

Parameter	Value
Framework	Darketnet
Image Dimensions	608×608 (LP-Net) and 384×192 (Char-Net)
Channels	3
Activation Function	Leaky RELU
Policy	steps
Epoch	5000
Max batches	6000 (LP-Net) and 54000 (Char-Net)
Batch size	64
Subdivision	16
Learning rate	0.001
Decay	0.0005
Momentum	0.9

and characters), the number of negative correctly detected (plates and characters), and the number of not detected (plates and characters), respectively. Object detection correctly depends on the Intersection over Union (IoU) value that measures the similarity between the ground truth, and the predicted bounding box. We choose IOU = 0.5. The average precision (AP) is described in Equation (8) as the area under precision and recall values. AP is the standard evaluation way for object detection networks.

$$AP = \int_{0}^{1} P_{(\mathrm{R})} dR \tag{8}$$

where $P_{(R)}$ is the precision over recall. The mAP value is the average of all classes of AP.

C. Results Analysis

We evaluate the Performance of LP-Net and Char-Net using the mean average precision with the threshold value = 0.5. After training both networks with 5k epoch. The mAP of Let-Net is 97.89 %, while the Char-Net Performance is 92.46% to predict the bounding boxes with the ground truth. Fig. 8 and 9 show the results obtained from the experiments on both networks. In Fig. 8, the model result is the bounding boxes for each LP and LP class inside the vehicle image, while in Fig. 9 the result is the bounding boxes for each character and its class inside the detected license plate.

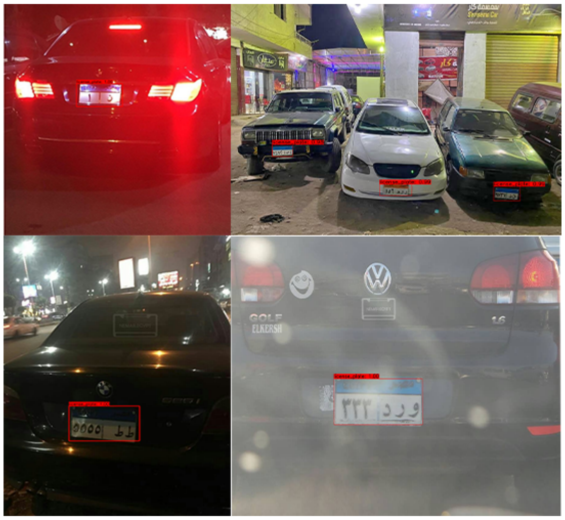


Fig. 8. LP-Net Results.

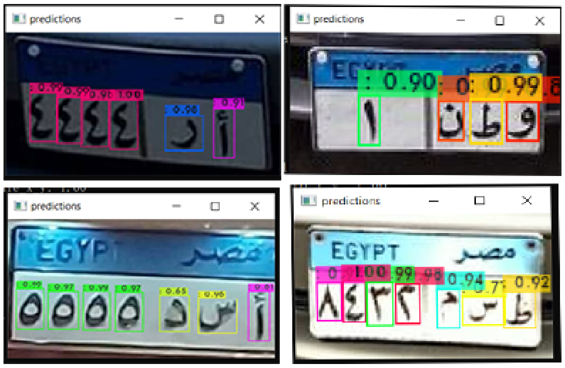


Fig. 9. Char-Net Results.

V. Conclusion

In this study, We present a real-time Egyptian license plate detection and recognition system. The proposed system is a pipeline of two deep convolutional neural networks. It can be deployed on GPU-less computing devices. The tiny-YOLOV3 model is the backbone for both networks. Also, We evaluated it on the publicly available Egyptian License plate (EALRP) dataset. The accuracy of the proposed system is 97.89% and 92.46% for license plate detection and character recognition, respectively. We suggest collecting more samples for the dataset with many variations and Implementing other YOLO model versions or Attention vision transformers to improve the accuracy of the proposed system.

References

[1] Y. M. Alginahi, “Automatic arabic license plate recognition,” *International Journal of Computer and Electrical Engineering*, pp. 454–460, 2011.

[2] S. M. Silva and C. R. Jung, “Real-time brazilian license plate detection and recognition using deep convolutional neural networks,” *Proceedings - 30th Conference on Graphics, Patterns and Images, SIBGRAPI 2017*, pp. 55–62, 11 2017.

[3] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Goncalves, W. R. Schwartz, and D. Menotti, “A robust real-time automatic license plate recognition based on the yolo detector,” vol. 2018-July, 2018.

[4] G. R. Gonçalves, D. Menotti, and W. R. Schwartz, “License plate recognition based on temporal redundancy,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pp. 2577–2582, 12 2016.

[5] H. H. Kim, J. K. Park, J. H. Oh, and D. J. Kang, “Multi-task convolutional neural network system for license plate recognition,” *International Journal of Control, Automation and Systems*, vol. 15, pp. 2942–2949, 12 2017.

[6] N. O. Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, “Deep learning vs. traditional computer vision,” 04 2019.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 779–788, 6 2015. [Online]. Available: https://arxiv.org/abs/1506.02640v5

[8] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” 4 2018. [Online]. Available: https://arxiv.org/abs/1804.02767v1

[9] ——, “Yolo9000: Better, faster, stronger,” *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 6517–6525, 12 2016. [Online]. Available: https://arxiv.org/abs/1612.08242v1

[10] S. M. Youssef and S. B. AbdelRahman, “Retracted: A smart access control using an efficient license plate location and recognition approach,” *Expert Systems with Applications*, vol. 34, pp. 256–265, 1 2008.

[11] S. Bayoumi, E. Korany, and S. Fathy, “License plate recognition system for egyptian car plates,” *International Conference on Image and Video Processing and Computer Vision*, 7 2010.

[12] Instagram, “cars_inegypt,” 2010, [Accessed 20-08-2020]. [Online]. Available: https://www.instagram.com/cars_inegypt

[13] ——, “exotics_in_egypt,” 2010, [Accessed 01-09-2020]. [Online]. Available: https://www.instagram.com/exotics_in_egypt

[14] ——, “nemar.egypt,” 2010, [Accessed 20-08-2020]. [Online]. Available: https://www.instagram.com/nemar.egypt

[15] ——, “egyptnumberplates,” 2010, [Accessed 01-09-2020]. [Online]. Available: https://www.instagram.com/egyptnumberplates/

[16] F. Marketplace, “Facebook marketplace egypt vehicles,” 2016, [Accessed 05-01-2020]. [Online]. Available: https://www.facebook.com/marketplace/category/vehicles

[17] Instaloader, “Instagram web scraping tool,” [Accessed 17-10-2021]. [Online]. Available: https://github.com/instaloader/instaloader

[18] Kiwibp, “Selenium script,” [Accessed 01-11-2021]. [Online]. Available: https://gist.github.com/Kiwibp/78cf224a0a5d0c2c33fdb371b8ebdb93#file-facebook-marketplace-selenium-py

[19] MOI, “Egyptian license plate specification,” 2008, [Accessed 12-03-2020]. [Online]. Available: https://traffic.moi.gov.eg/English/OurServices/InfoServices/InteriorMinisterDecision/Pages/license-plates-taxes-traffic-fees.aspx

[20] D. Sun, “Ybat - yolo bbox annotation tool,” [Accessed 04-07-2020]. [Online]. Available: https://github.com/drainingsun/ybat

[21] OpenCV, “Opencv,” [Accessed 16-10-2021]. [Online]. Available: https://opencv.org/

[22] AlexeyAB, “Yolov3,” [Accessed 16-10-2021]. [Online]. Available: https://github.com/AlexeyAB/darknet

[23] J. Davis and M. Goadrich, “The relationship between precision-recall and roc curves,” *ACM International Conference Proceeding Series*, vol. 148, pp. 233–240, 2006.

Building an Arabic Dialectal Diagnostic Dataset for Healthcare

Jinane Mounsef
Department of Electrical Engineering and Computing Sciences
Rochester Institute of Technology
Dubai, United Arab Emirates

Maheen Hasib
School of Mathematical and Computer Sciences
Heriot Watt University
Dubai, United Arab Emirates

Ali Raza
Department of Electrical Engineering and Computing Sciences
Rochester Institute of Technology
Dubai, United Arab Emirates

Abstract—**Accurate diagnosis of patient conditions becomes challenging for medical practitioners in urban metropolitan cities. A variety of languages and spoken dialects impedes the diagnosis achieved through the exploratory journey a medical practitioner and patient go through. Natural language processing has been used in well-known applications, such as Google Translate, as a solution to reduce language barriers. Languages typically encountered in these applications provide the most commonly known, used or standardized dialect. The Arabic language can benefit from the common dialect, which is available in such applications. However, given the diversity of dialects in Arabic in the healthcare domain, there is a risk associated with incorrect interpretation of a dialect, which can impact the diagnosis or treatment of patients. Arabic language dialect corpuses published in recent research work can be applied to rule-based natural language applications. Our study aims to develop an approach to support medical practitioners by ensuring that the diagnosis is not impeded based on the misinterpretation of patient responses. Our initial approach reported in this work adopts the methods used by practitioners in the diagnosis carried out within the scope of the Emirati and Egyptian Arabic dialects. In this paper, we develop and provide a public Arabic Dialect Dataset (ADD), which is a corpus of audio samples related to healthcare. In order to train machine learning models, the dataset development is designed with multi-class labelling. Our work indicates that there is a clear risk of bias in datasets, which may come about when a large number of classes do not have enough training samples. Our crowd sourcing solution presented in this work may be an approach to overcome the sourcing of audio samples. Models trained with this dataset may be used to support the diagnosis made by medical practitioners.**

Keywords*—*Dialectal Arabic (DA); healthcare diagnosis; natural language processing (NLP); multi-class labeling; crowd sourcing

I. Introduction

Research into healthcare practices often highlight language barriers as a major hurdle that impedes a seamless doctor patient interaction and hinders the possibility of positive diagnostic outcomes [1], [2], [3]. In emergency rooms, where timing is critical and information exchange between the patient and the doctor can save lives, the language barrier is seen as an obstacle that must be overcome. It is important to understand the role that language plays, as there has been an increase in culturally and linguistically diverse populations among patients and practitioners [2]. As a result of this ongoing global migration and globalization, many societies all over the world have become multicultural, with many migrants who do not speak the official language of the host country. In the United Arab Emirates (UAE) in particular, the local citizens make up only about 11% of the population, with the rest being from various parts of the world, mostly from Southeast Asian countries and various Arab countries [4]. With this diversity in culture and population, it causes a communication gap in the healthcare service and creates challenges in providing quality individual and holistic healthcare [5].

A large number of published studies [6], [7] address the language barrier that exists between the doctor and the patient, resulting in a poor understanding of diagnosis, relevant investigations, and medication instructions [7]. Surveys, such as that conducted by [8], have shown that language barrier can also lead to poor comprehension and compliance with recommendations for follow-up and treatment. This increases the likelihood of the occurrence of adverse medical events, thereby lowering patient satisfaction. Hence, communication barriers must be considered as part of any strategy aimed at improving patient safety and risk management in healthcare organizations [9].

Attempts to solve this communication problem have been reported in [10], [11], [12], [13] by using Google Translate (GT) or an interpreter. However, when language is a barrier, GT remains the most easily accessible and a free initial mode of communication between the doctor and the patient [10]. A good medical interpreter needs to be able to accurately translate the complex medical terminology. Evidence in [14] suggests that when limited English proficiency patients have access to skilled professional interpreters or bilingual physicians, they have better communication, patient satisfaction, and outcomes, as well as fewer interpretation errors.

The limitations related to healthcare and diagnosis have been reported in [14], [15], [16]. GT has only 57.7% accuracy and should not be relied on for critical medical communications [10]. In [12], it is mentioned that those patients who require interpreters but do not receive them, have a poor self-reported understanding of their diagnosis and treatment. Ad hoc interpreters misinterpret or omit up to half of all physicians' questions and are more likely to make clinically significant errors. The three most commonly used forms of interpreters are ad hoc, professional, and telephone interpreters. For ad hoc interpreters, people commonly use friends or family members of the patient or staff in the work setting, such as housekeepers, secretaries, or medical personnel, who are untrained in interpreting. It is increasingly recognized that the use of untrained or ad hoc interpreters can lead to inaccurate communication and ethical breaches.

Language barrier leads to a communication breakdown between the patient and healthcare providers [6], [17]. Hence, it is critical for the doctor and the patient to communicate effectively, which is why it is advantageous if the doctor is fluent in the patient's language [18], [19]. However, doctors may not always be able to communicate in the patients' native language, which creates hurdles, as shown in [20].

Early health communication linguistic research [21] focused mainly on small datasets, such as language samples gathered from face-to-face clinical interactions or research interviews [22]. The main drawback of this was that the findings presented were based on limited datasets and hence, researchers have started to use corpus linguistic methods [23]. A corpus is a collection of authentic text or audio organized into datasets [24].

The Arabic language is classified into three forms: classical, modern standard and dialectal. Classical Arabic (CA) is the language used in both ancient history literary and religious texts. Modern Standard Arabic (MSA) is the "official" Arabic used in books, magazines, media, legal documents, etc. Dialectal Arabic (DA) is the informal Arabic that native speakers use to communicate in their daily lives. Research in [25], [26], [27], [28] focused on the development of an Arabic language corpus, such as Egyptian, Tunisian and Iraqi dialects.

Over the last decade, Arabic and its dialects have made growth in the field of Natural Language Processing (NLP) [29]. Dialects vary and mutate over the large geographical span of the Arab peninsula with popularity borne by the Egyptian dialect in a majority of countries in the Middle East and North African regions. The UAE population distribution provides the motivation for the work reported in this paper. Of the 40% Arab population in the UAE, Emirati and non-Emirati Arabic groups are almost in equal proportion. This allows us to focus our research efforts on the Emirati and Egyptian dialects based on the fact that the Egyptian dialect is the most spoken in the non-Emirati Arab community. Although research in [30] focused on developing a pediatric assistant that supported Arabic dialects, particularly Egyptian dialects, there is a lack of research for both Emirati and Egyptian dialects in the UAE healthcare sector.

With the growth of AI applications, researchers are looking for ways to use NLP (a strand of AI) to solve real-world problems related to language. So, a dataset that can enhance the functionality of NLP-based applications is highly desirable for researchers in this field and is a main contribution for science. Researchers working on language translation tools and chatbots will be very interested in such tools when focusing on the development of solutions for the healthcare industry. These datasets can be used to develop machine learning models to detect and classify linguistics.

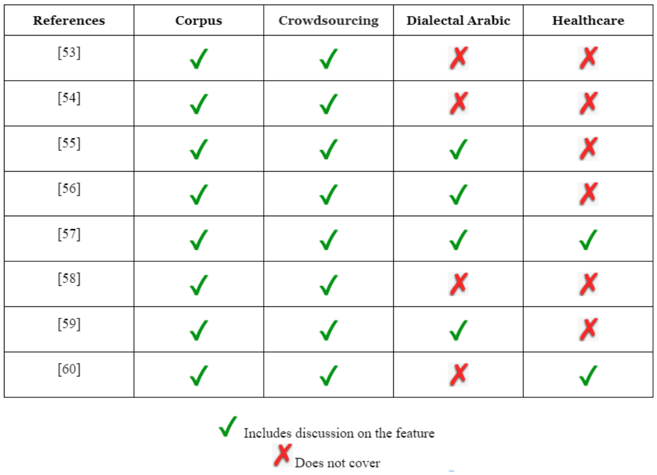
In this work, the approach used by medical practitioners in the UAE were analyzed following some interviews with doctors representing the Dubai Health Authority (DHA). A database of questions was developed in the English Language and a platform was developed to allow users to record a translated version of the typical response to a question in their Arabic dialect (Emirati or Egyptian). These responses were labeled and stored as an audio dataset in a raw mp3 format for researchers to use. Recordings with poor audio quality were manually removed after an inspection. The development of AI-based tools can improve the diagnostic outcomes for patients and lead to a better care. Issues that are diagnosed early allow for treatments, which can lead to correct prognosis and timely interventions.

The remainder of this paper is organized as follows: Section II provides the related work in the field of Arabic NLP (ANLP); Section III discusses the DA challenges in healthcare; Section IV presents our approach and methodology for building a dataset to train a Machine Learning (ML) algorithm; Section V deals with the ML approach to which the dataset can be applied. Finally, Section VI provides the limitations of the work in addition to the social and ethical considerations. We conclude this paper with future directions of research that can be supported through this dataset and its enhancement.

II. Literature Review

A considerable amount of literature has been published on the growth of the DA, which has been recognized as the primary language for informal communication [31]. This growth has been aligned with the emerging trends in online social media platforms (e.g. Twitter, Snapchat, Facebook etc.). The online social media factor attributes to the varsity of DA, which has led to a wider use in offline social interaction. As highlighted in [32], the forms of DA differ depending on the geographic distribution and the socio-economic conditions of the speakers. Each form of DA is considered a divergence from the formal variety, the MSA. This divergence is evidenced in the challenges presented by [33], highlighting Arabic Internet users adopting a mixture of MSA and DA, thereby increasing the challenges related to text processing for machine translation through NLP.

The recent advances in computational power and processing of large data arrays have led to the enhancement of NLP as a technique to empower machines in gaining a better understanding of the human language [34]. Deep learning (DL), a consequence of the rise in computational power, has increased the opportunity for more tasks to be carried out by NLP. The Arabic NLP community (ANLP) [35] has yet to grasp the advantages offered by NLP with DL, despite the growth of DA aligned with social media platforms mentioned earlier. However, recent work published by [36] shows a new focus on ML and DL, based on lexicon and corpus approaches. Islam et al. [37] discuss the importance of NLP in the healthcare industry with researchers finding ways to improve diagnostics. Furthermore, automation for initial and informal diagnosis has been envisaged through the use of NLP to provide an on-demand self-service for patients [38]. The communication between a doctor and a patient usually involves personal questions, which reveal intimate information necessary for an accurate diagnosis [19]. These challenges are amplified in the Arabic culture and the framework of respect and politeness around the spoken dialogue [39]. To cater for these challenges, techniques used by practitioners have included the use of an interpreter [40]. However, the framework of politeness mentioned earlier also reduces the success of using an interpreter in the Arabic culture. Another



References	Corpus	Crowdsourcing	Dialectal Arabic	Healthcare
[53]	✓	✓	✗	✗
[54]	✓	✓	✗	✗
[55]	✓	✓	✓	✗
[56]	✓	✓	✓	✗
[57]	✓	✓	✓	✓
[58]	✓	✓	✗	✗
[59]	✓	✓	✓	✗
[60]	✓	✓	✗	✓

✓ Includes discussion on the feature

✗ Does not cover

Fig. 1. List of Published Papers using Crowd Sourcing.

alternative approach used by doctors is known as code switching [41], where the bilingual capability of the practitioner can be used to combine English words and terms (of important medical consequence) and words in DA, which can provide an atmosphere of comfort and respect for the patient in the cultural setting. Although code switching has been reported to be a positive and useful approach, the bilingual capability is not ubiquitous, given the divergence of DA. Hence, approaches, which are supported by artificial intelligence (AI), play a role in bridging the gap. This gap is wider, where the growth in healthcare tourism [42] is supported by the deployment of practitioners accustomed to MSA or a different strand of DA [43].

Approaches in sentiment analysis, which apply NLP as reported in [44], have been used in healthcare based on text computation extracted from web sources, such as blogs and social media, and with no methodical approach used for the attainment. Again, the politeness of the Arabic culture plays a significant role in the level of accuracy that may be gained through these approaches, as there is a limited directness. Furthermore, online sources will also suffer from a lack of contextual information, which can be used to train models used in AI. Researches [31], [44], [45], [46] were developed using a corpus-based approach for sentiment analysis in healthcare. The techniques that have been used are Lexicon. There are several challenges mentioned in [45], [46] facing the sentiment analysis and evaluation process, especially in the medical domain. There are health related blogs and forums where people discuss their health issues, symptoms, diseases, medication, etc. [47]. These challenges become obstacles in analyzing the accurate meaning of sentiments and detecting the suitable sentiment polarity. The approaches used to overcome these challenges are discussed in [48].

ML implementations of audio analytics have gained popularity in applications, such as Alexa, Siri, Google Assistant and Google Home. The capabilities offered are founded on models, which can extract information from audio signals [49]. Audio processing capabilities supported by open-source codes and public big data have led to an accelerated development of applications, such as Shazam, which provides an audio similarity search capability [50]. Applications of audio analytics can range from customer sentiment analysis in recordings taken from call centers to media content monitoring for parental control. In the domain of healthcare, several ML-based approaches have been proposed to consider clinical decisions supported using text-free data [51]. In recent years, clinical speech and audio processing have offered new opportunities, such as automatic transcription of patient conversations, synthesis of clinical notes, as well as identification of disorders related to speech [52]. Chatbots, such as Babylon Health, have been developed for diagnosis and prevention of diseases. Such chatbots use speech recognition technologies to compare symptoms reported by the user with a database of diseases to recommend a course of actions based on the identified disease in addition to the patient history. In the special case of machine translation, there is currently limited evidence in the literature of the use of healthcare-oriented ML-based translators in clinical practice [53]. Instead, physicians commonly use either professional interpreters, who are usually costly, or digital translating services, such as GT, which are not tailored to properly function with a medical lexicon.

There are different techniques for processing the audio files for ML. The approaches used in the literature are audio spectrograms in [54] and Mel Frequency Cepstral Coefficients (MFCC) in [55]. In [56], the input of fixed length video segments of 10 seconds was converted to an audio spectrogram and fed into a Convolutional Neural Network (CNN) based on Alexnet [57] and VGG-16 [58] architectures. In [55], the use of audio spectrograms to feed into a CNN with the use of MFCCs was discussed. The authors also used a second approach, where the audio spectrogram was used as an input, using a modified VGG-16 architecture based on transfer learning [59]. Results in [55] showed that the accuracy using MFCC was higher than the spectrogram-based approach.

The approach proposed in this paper supports speech recognition (SR) under text and speech processing as a branch of NLP. Advances in SR through big data and DL have yielded significant gains through innovative approaches, as found in recent academic work [60], [61]. Limited work on the use of an audio corpus is referenced in [54], where voice recognition based on whole sentences without speech segmentation is done. To improve and innovate in the field of voice recognition for healthcare diagnosis in the DA, this corpus provides the big data component, which is labelled and can be used for supervised learning. The labelling approach of the corpus can support the development of voice recognition with DL for improvement in the healthcare diagnostic outcomes. The approach used to create the corpus is similar to Amazon Mechanical Turk. Other approaches, which use the crowd sourcing and labelling mechanism, are mentioned in Fig. 1.

III. Healthcare Challenges with DA

With the discovery of oil in the UAE in the late 1950s, a huge economic and social transformation began resulting in an increase of the need for foreign labor. Since then, the UAE has seen a huge number of expatriates trickling in to gain from the profitable economic bustle including trade, real estate, construction, and healthcare, thereby outnumbering the national population. While the nationals whose spoken language is Arabic only amount to 11.48% of the entire

TABLE I. SAME MEANING OF DIFFERENT HEALTHCARE WORDS IN EGYPTIAN AND EMIRATI DIALECTS

Egyptian Dialect	Emirati Dialect	Meaning
وجع	عوار	pain
غمَمان	لوعة	nausea
أرّجع	بزوع	I vomit
بتوجعني	يعورني	It hurts

TABLE II. EXAMPLES OF HEALTHCARE STATEMENTS IN DIFFERENT DAS WITH THEIR TRANSLATIONS IN ENGLISH

Questions/ Answers	Translation by GT	Correct Translation
زوري واجعني (Egyptian dialect)	visit and see me	my throat hurts
انا دايخ (Egyptian dialect)	I am deaf	I am dizzy
انا عيان (Egyptian dialect)	I am an eye	I am sick
شو رح يصير لو ما سويت هاي العملية؟ (Syrian dialect)	What would happen if you did not make this process happen?	What will happen if I do not get the surgery?
حاسة حالي تعبانة (Jordanian dialect)	My current sense is tired	I am tired
فيني لوعة (Emirati dialect)	Vinny is crazy	I feel nauseated

population in 2022, the expatriates total around 88.52% of the population, mainly coming from India, Pakistan, Bangladesh, Philippines, Iran, Egypt, Nepal, Sri Lanka and China [4]. As most of these migrants are non-Arabic speakers, the common communication language used in the different public and private service sectors in the UAE is English. The Arabic language has also been a barrier to Arabic speakers across the country, as the Arab migrants use different dialects or DAs, which diverge from the MSA, and are often classified regionally as Egyptian, North African, Levantine, Gulf, and Yemeni or sub-regionally as Tunisian, Algerian, Moroccan, Lebanese, Syrian, Jordanian, Saudi, Kuwaiti and Qatari [62]. Moreover, many of the Arabic speakers cannot use the English language to communicate efficiently with non-Arabic speakers. This suggests that the language in the UAE is a contributing factor to misinformation in several critical service providers including the healthcare sector.

In the UAE, the two mostly spoken Arabic dialects are Emirati (11.48%) and Egyptian (4.23%) [4]. While the majority of the Arabic speaking residents use these two dialects to communicate on a daily basis [63], hospitals are staffed mostly with physicians who speak exclusively English at workplace. Therefore, these physicians are frequently obliged to deal with Arabic speaking patients who have no language in common. Moreover, the diversity of the Arabic language through its several DAs poses another serious challenge that needs to be addressed [62]. Table I shows an example of several Arabic healthcare terms expressed differently in Emirati and Egyptian dialects, but having the same meaning. A situation of this kind, with barriers in language and medical understanding, creates serious problems for quality, security and equitability of medical care. Rosse et al. [64] defined a language barrier as “a communication barrier resulting from the parties concerned speaking different languages” and supported previous claims that these barriers pose a significant threat to quality of care and patient safety.

Therefore, the language barrier poses challenges for achieving high levels of satisfaction among physicians and patients, providing high-quality healthcare service, and maintaining patient safety. To address these challenges, several solutions are available today, but they all have their drawbacks. There has been recently a rising demand of English-Arabic translating services for medical, anatomy and healthcare lexicons in Arabic cosmopolitan countries, which usually attract multilingual/multidialectal expatriates [65]. Knowing that professional interpreters are very expensive and are not always available for some dialects, the framework of politeness in the Arab culture also reduces the success of using an interpreter for DA speakers. On the other hand, GT, increasingly often used when no other alternatives exist, is known to be unreliable for medical communication [66] and human interference is greatly needed to produce accurate and effective translation. For instance, an Egyptian patient reporting a pain in the leg, ””رجلي وجعاني , is interpreted as “My feet are hungry” instead of “My leg hurts”. Another example shows the inadequacy of GT, which fails to translate two different statements having the same meaning into the same expression. The two Egyptian dialect statements ””انا مصدع and ””انا عندي صداع ,which both report a headache, are construed as “I’m cracked” and “I have a headache”, respectively. Table II, which illustrates some examples of translation from several DAs to English, shows the inaccurate translation of GT when used in the healthcare sector.

All these factors lead us to explore and suggest a better solution to the language dilemma that physicians face in providing high-quality and safe care to DA-speaking patients with limited English proficiency in the UAE. Our work proposes a new approach of efficiently addressing this problem by using the crowd sourcing and labelling mechanism for the Emirati and Egyptian dialects in the medical diagnosis context. To our knowledge, no other work has followed this approach to address the challenge of English-Arabic interpretation in the medical field, specifically related to the emergency diagnosis questionnaire.

IV. APPROACH AND METHODOLOGY

The crowd sourcing approach used in this work is termed as “crowd labeling”. In this approach, an automated labeling implementation is developed, which allows for audio recordings to be labeled at the time they are recorded and uploaded by contributors.

A. Architecture Overview

Fig. 2 provides an overview of the application architecture using HTML, CSS, JavaScript, PHP and Google Developer APIs. Visitors to the webpage can see a landing page, which offers them the option to record their phrase in the Emirati or Egyptian dialect. The selection prompts a back-end process on the web server to get a random question and associated phrase from a pool, which is hosted in a Google sheets document. This is done using Google developer APIs.

The contributor records the phrase and uploads it to the audio sample database, which is hosted in Google drive. This repository is available for public access for other researchers. The audio files are labelled automatically. When a file is

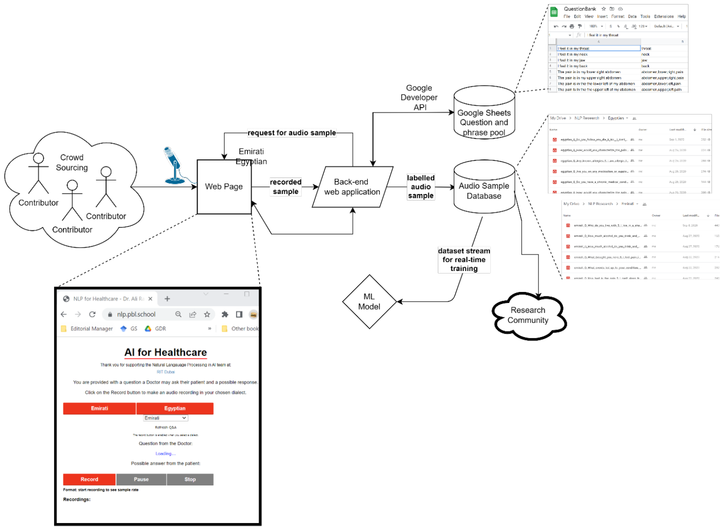


Fig. 2. High-Level Implementation Architecture.

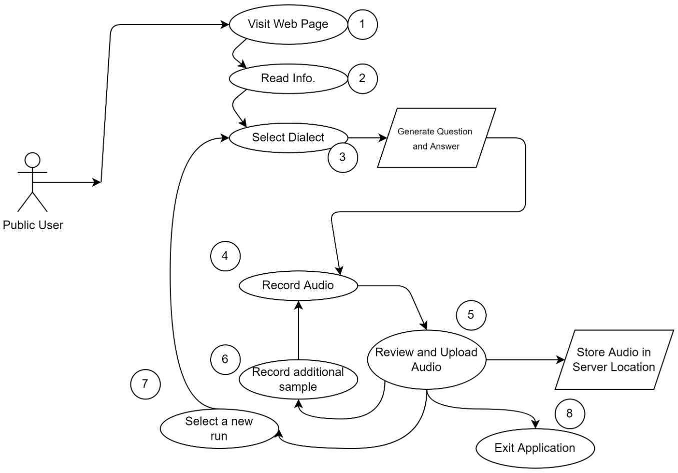


Fig. 3. Crowd Labeling Application Flow Diagram.

uploaded, the following attributes are included in the file name: <dialect>_Q_<possible question from Doctor> _S_<possible answer from patient>_L_ <predefined labels>.mp3

An example from the repository is provided here: emirati_Q_What_brought_you_here_S_I_feel_ pain_in_my_leg_L_leg_pain.mp3

The audio samples are recorded with a 48 MHz sampling rate and stored in the mp3 format.

B. Application Overview

Fig. 3 provides an overview of the implementation with a summary of the steps:

1) User visits the webpage (https://nlp.pbl.school/)
2) User reads the welcome note.
3) User selects the dialect they wish to contribute in.
4) Once the question-and-answer set is generated, the user records the audio.
5) The user reviews the recording quality and uploads it to the Google drive storage location. The user can proceed to step 7 to continue with contributions or proceed to step 8 and exit the application by closing their browser page.
6) If the audio sample is not clear, the user can decide to re-record it and return to step 5.
7) The user can choose to launch a new contribution and restart at step 3.
8) The user can choose to exit the application at any time.

C. Access to Repository

The repository of the collected audio files can be accessed from Google drive using the link below. As new recordings are published, this drive is automatically updated. Researchers have unlimited access to the ADD dataset:

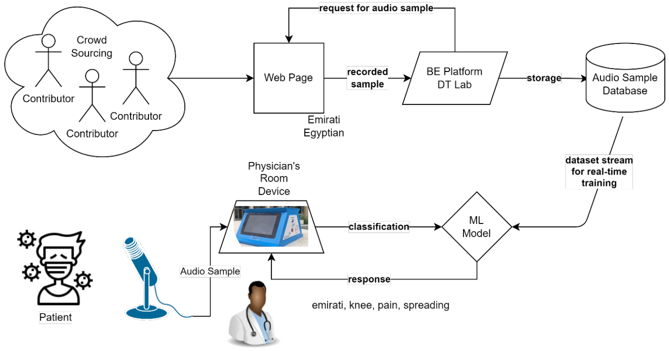


Fig. 4. Implementation of ML Pipeline for Patient-Doctor Diagnostic Device.

https://drive.google.com/drive/folders/160HE3q_ FcqNJMyC4M5Hx1e2SffKhDeW8?usp=sharing

Access to this repository can also be automated using Google drive APIs by researchers who intend to create a continuous ML pipeline. With this approach, each time a new file is uploaded, the ML pipeline implemented can fetch the file and update the ML algorithm. The current implementation of this is shown in Fig. 4. The implementation consists of a two-monitor device based on the Raspberry Pi 4, and placed in the room of a physician. The doctor can select a question, which is played in the dialect of the patient. The response from the patient is recorded and uploaded to the ML model for classification. The response from the ML model provides the physician with the phrase from the database that best matches the input from the patient.

Preliminary results on the accuracy of the classification have been produced, as described in Section V. We intend to publish detailed results in the future when the size of the corpus reaches a minimum threshold.

V. ML Proposed Applications

We address in this section the need for capacity development in this area by providing some conceptual ML methods

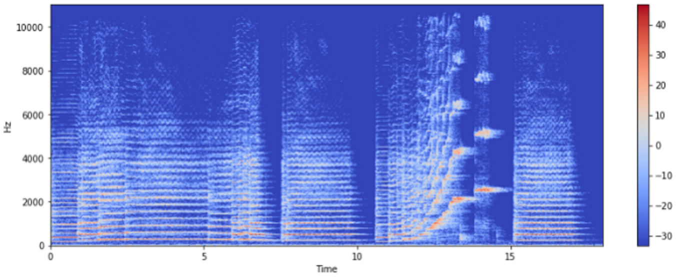


Fig. 5. Spectrogram of an Audio Clip with a Spectral Range of 0 kHz to 10 kHz.

for researchers to developing predictive algorithms for several healthcare speech recognition-related applications. Examples of such applications include speech tagging, dialect classification, diagnosis multi-labelling and language translation, using our freely available public ADD dataset for Emirati and Egyptian dialect lexicons.

Libraries, such as Librosa, provide useful tools for audio signal processing, using a Fourier Transform to convert an audio signal into a frequency domain from the time domain. Librosa supports the display, processing, and analysis of key spectral features, such as spectral-flux, spectral centroids and fundamental frequency components, which are essential features for ML-based spectral analysis [49]. Fig. 5 shows an implementation of Librosa's spectral display capabilities in which spectral images of our collected audio samples are generated. The image is normalized to a specific color profile and a frequency profile from 0 to 4 kHz, which aligns to the human speech audio range.

In a typical classification application, the images can be processed and numerically represented with the pixel color information being the input for each neuron in an Artificial Neural Network (ANN) or Convolutional Neural Network (CNN) model, such as AlexNet or GoogLeNet, as proposed by Boddapati et al. [67]. The training of the model will allow for classification of dialects and diagnosis types. This is summarized by the process pipeline shown in Fig. 6. In this section, we develop and test the proposed design process on the ADD dataset, as one possible approach of using the dataset for Arabic/English translation of the emergency diagnosis patient answers.

A. Audio Processing

A total of 301 recordings in the Egyptian dialect and 138 recordings in the Emirati dialect are collected as part of the ADD after manually removing the faulty audio samples from 12% of the Egyptian dialect recordings and 22% of the Emirati dialect recordings. The errors include wrong translation, blank recordings, and duplications. The maximum duration for an audio sample is 10 seconds, the minimum is 1 second, and the average is between 2 and 3 seconds.

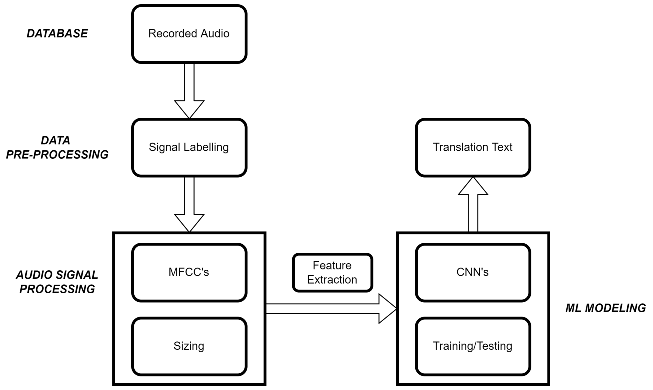


Fig. 6. Proposed Classification Pipeline.

Next, we consider the recorded raw audio samples, where a sample of an Egyptian dialect speech waveform is shown in Fig. 7. The waveform only shows the change of the signal's amplitude over time without giving any insight about the different frequency components pertaining to the recorded audio signal. Therefore, we convert the audio waveform to a spectrogram, which is a 2D image representing sequences of spectra with time along one axis, frequency along the other, and brightness or color indicating the strength of a frequency component at each time frame. This representation can thus be used with CNNs that are usually applied on images and can be applied directly to sound in this case. Moreover, learning more about the signal's frequencies gives us a better understanding of the recorded audio signal and allows us to filter out any unwanted disturbance, as distortion and noise can be visualized in a spectrogram.

B. Machine Learning Models

For speech classification, Mel Frequency Cepstral Coefficients (MFCCs) are commonly used for their classification and identification effectiveness, as they can describe concisely the shape of the spectral envelope expressed on a Mel-scale. However, MFCCs are also known as being "lossy" representations, which are preferably ruled out when working with a high-quality sound. Therefore, the spectrograms can be used directly as 2D images to feed pre-trained CNNs. In the case of a limited size dataset, transfer learning is used to relax the hypothesis that the training data must be large, independent and identically distributed with the test data. This motivates many work [68], [69], [70], [71], [72], [73] to use transfer learning in the presence of insufficient training data for speech and language classification. A network, which is pre-trained on a large-sized dataset, such as the ImageNet [74], will keep its structure and connection parameters, when used by a network-based deep transfer learning, to compute intermediate image representations for smaller-sized datasets. Therefore, a network, such as a CNN, is trained on the ImageNet to learn image representations that can be efficiently transferred to other visual recognition tasks with limited amount of training data. The front-layers of the network are operated as a feature extractor, where the extracted features are classified using ML classifiers, such as a support vector machine (SVM).

C. Preliminary Results

We propose a similar machine learning model for the spectrograms classification to identify the English translation of the corresponding Arabic audio sample. Fig. 8 shows the different steps followed in the described model. We perform

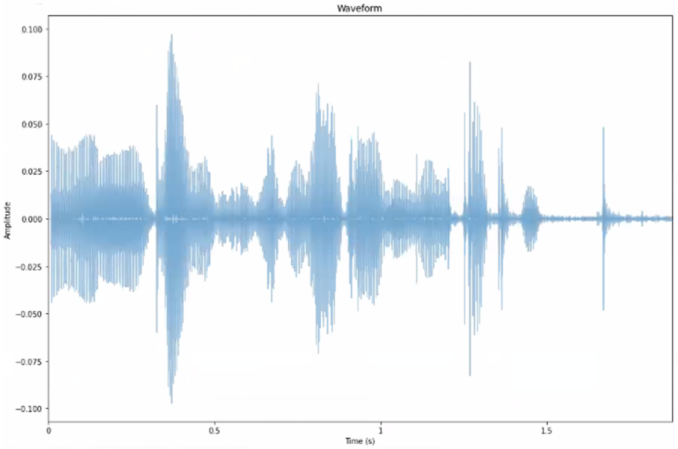


Fig. 7. Audio Waveform Sample of an Egyptian Dialect Recorded Answer.

identification experiments with a closed-set experimental protocol, where our model should predict the label to which the input image belongs. For experimentation, we use four classes of the dataset labeled as: "I smoke", "I do not smoke", "I live with my family", and "I live alone". The training set consists of 80% of the spectrogram images, while the testing set includes the remaining 20% of the images. Since the ADD dataset is relatively small, it is easy for the CNN model to over-fit and not generalize well on the testing data. To alleviate this problem, we augment the dataset with a factor of five by editing the pitch of the audio recordings to four different values. Other augmentation techniques are possible to use, such as time stretching or adding noise. We use the augmented training set of the resulting spectrograms on a VGG16 network. We extract features from different layers of the same network to explore the different classification accuracies. We perform an exhaustive search on the layers and report results with the layer that gives the highest accuracy. We find that the best performance corresponds to the last convolutional layer of VGG16. We use the extracted features to train a one-against-one multi-class linear SVM. The achieved recognition accuracy is of 78%.

Although the result might not look promising enough, it shows that the proposed model performs decently well for a small-sized dataset and proves to utilize the ADD in one of many ways to implement Emirati and Egyptian dialect sentences classification into their respective English translated expressions.

VI. Limitations, Social and Ethical Considerations

The main limitations of the approach presented in this work were the residual response bias of participants, the short period of time to conduct the research and failure to fully explore all the possible responses for both dialects.

The current work relies on translation crowdsourcing, which is particularly known for its innovative approach, combining the concept of crowdsourcing and that of natural and non-professional translation. The translation project that formed the basis of this work comprised of untrained translator

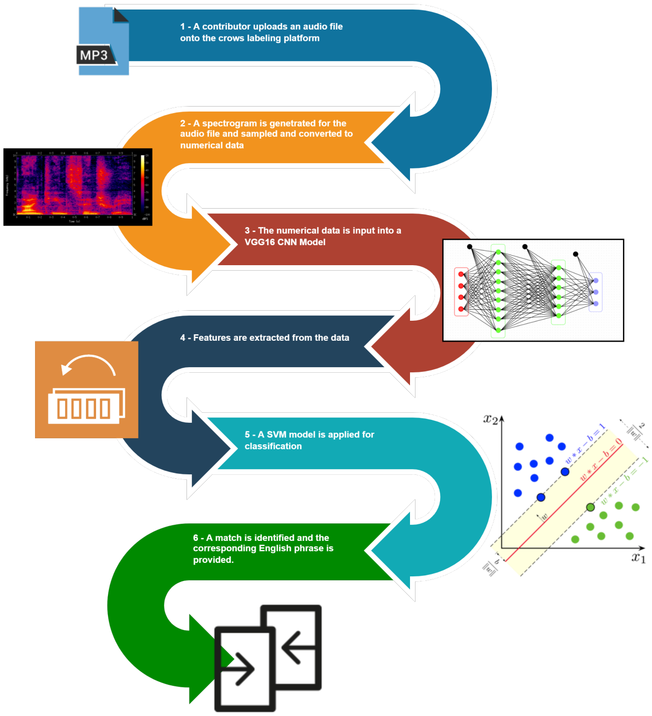

MP3
1 - A contributor uploads an audio file onto the crows labeling platform
2 - A spectrogram is genetrated for the audio file and sampled and converted to numerical data
3 - The numerical data is input into a VGG16 CNN Model
4 - Features are extracted from the data
5 - A SVM model is applied for classification
6 - A match is identified and the corresponding English phrase is provided.
x_2
x_1
$w * x - b = 1$
$w * x - b = 0$
$w * x - b = -1$
w
$\frac{2}{\|w\|}$
$\frac{b}{\|w\|}$

Fig. 8. One Proposed Design of an ML Scheme Application.

participants who were native speakers of the Arabic dialect with no previous experience of translation. Even though their contribution to the crowdsourcing platform, as native speakers, undoubtedly gave a high credibility to the correctness of the translated responses, there was still an obvious response bias in the research given that they were translator novices, with many participants translating the response differently based on their knowledge and background. The work shows that these limitations in volunteer translation suggest that one might bridge the gap between trained and untrained translators through collaboration and mutual training in this kind of projects as in [75].

To accommodate building the dataset and making it public to the community, only a short period of time was available for the research. The crowdsourcing platform and initial recordings were achieved in nine months. Participants were selected from the close network of school, family, and friends who spoke the Emirati and Egyptian dialects. Time restrictions limited the size of the dataset, which includes a total of 301 Egyptian and 138 Emirati dialects recordings. This resulted in an imbalanced dataset on multiple levels. The Egyptian dialect recordings are almost double the size of the Emirati ones, knowing that the acquaintances of the crowdsourcing developers consisted mainly of Egyptians. On the other hand, by looking at the participants gender figures, 95% of the Emirati participants are male, while 63% of the Egyptian participants are female. Finally, most of the participants (80%) are from the young generation, thereby minimizing the contribution of children, middle age and elder samples of the Emirati and Egyptian populations. This systemic inequity based on age and gender disparities has received a lot of attention recently in voice biometrics [76] and reveals that age and gender subgroups have a significant different voice characteristic. In fact, non-negligible gender disparities exist in speaker identification accuracy and show that the average accuracy can be significantly higher for female speakers than males due to mainly voice inherent characteristic difference [76]. The results in [77] also indicate a significant age effect on the voice acoustic parameters (fricative spectral center of gravity, spectral skewness, and speaking STSD) revealing that certain speech and voice features change with age.

Even though the research was deemed successful by curating the first medical diagnosis dataset for the Emirati and Egyptian dialects, it did fail to completely explore all the possible responses to the emergency diagnosis questionnaire, as not all the questions were presented to the participants in the crowdsourcing platform, and thus many responses missed being recorded in the Arabic dialects, at least in either one of them. In hindsight, participants should have been provided with a wide palette of different responses for them to translate before they were able to stop recording.

In order to protect the rights of the research participants and also to maintain scientific integrity, it was important to take ethical and social considerations into account [78]. The participants were therefore given the option of voluntary involvement, which allowed them to withdraw at any point without any obligations. It was also made clear to them that there were no negative consequences or repercussions to their refusal to participate. The identities of the research participants were kept anonymous, as no personal identifiable data were asked, such as their names, email addresses, phone numbers, photos and videos. The research used data pseudonymization, where the audio recordings of the participants were given artificial identifiers, such as sample001, sample002 etc.

VII. Conclusion

In this work, the implementation of a crowd labeling approach to develop an audio corpus is proposed. An application of the audio corpus to train an ML algorithm, which interfaces with a device in a physician's room is provided. Although results on the accuracy of the classification of this application are preliminary and incomplete due to the small size of the dataset, the implementation approach can be replicated by researchers by splitting the corpus into training and testing sets to evaluate different ML techniques or test a specific ML pipeline. The crowd labeling approach developed here can be extended to other applications, where data is collected from public contributors. Special care has been taken to ensure that no personal identifiable information about the contributors is collected or stored. The future proposed work will collect more audio samples and include additional Arabic dialects. A dashboard will be supporting the dataset to provide statistics on the collected audio samples.

Acknowledgment

The authors would like to thank Hana Abbas, Anan Hefny, Haneen Al Attar, Mohamed Dorrah, Amr Mohamed, Arjun Nair, Habiba Elshahat, and Sherif Reda for helping with this work.

References

[1] K. Gerrish, R. Chau, A. Sobowale, and E. Birks, "Bridging the Language Barrier: the Use of Interpreters in Primary Care Nursing," *Health & Social Care in the Community,* 2004, 12(5), 407-413.

[2] R. F.I. Meuter, C. Gallois, N. S. Segalowitz, A. G. Ryder, and J. Hocking, "Overcoming Language Barriers in Healthcare: a Protocol for Investigating Safe and Effective Communication when Patients or Clinicians use a Second Language," *BMC Health Services Research,* 2015, 15(1), 1-5.

[3] P. A. Ali and R. Watson, "Language Barriers and their Impact on Provision of Care to Patients with Limited English Proficiency: Nurses' Perspectives," *Journal of Clinical Nursing,* 2018, 27(5-6), e1152-e1160.

[4] GMI. United Arab Emirates Population Statistics 2022, 2022, https://www.globalmediainsight.com/blog/uae-population-statistics/. Accessed June 5, 2022.

[5] E. Hadziabdic and H. Katarina, "Working with Interpreters: Practical Advice for use of an Interpreter in Healthcare," *International Journal of Evidence-Based Healthcare,* 2013, 11(1), 69-76.

[6] H. S. Al-Neyadi, A. Salam, and M. Malik, "Measuring patient's satisfaction of healthcare services in the UAE hospitals: Using SERVQUAL." *International Journal of Healthcare Management,* 2018, 11(2), 96-105.

[7] A. Saqer and Q. Alaa, "Language Miscommunication in the Healthcare Sector: A Case Report." *Journal of Patient Safety and Quality Improvement,* 2019, 7(1), 33-35.

[8] C. L. Timmins, "The impact of language barriers on the health care of Latinos in the United States: a review of the literature and guidelines for practice," *Journal of midwifery and women's health,* 2002, 47(2), 80-96.

[9] T. Loney, T. C. Aw, D. G. Handysides, R. Ali, I. Blair, and M. Grivna, "An analysis of the health status of the United Arab Emirates: the 'Big 4' public health issues," *Global Health Action,* 2013, 6(1), 20100.

[10] S. Patil and P. Davies, "Use of Google Translate in Medical Communication: Evaluation of Accuracy," *British Medical Journal,* 2014, 349.

[11] E. Hadziabdic and H. Katarina Hjelm, "Arabic-Speaking Migrants' Experiences of the use of Interpreters in Healthcare: a Qualitative Explorative Study," *International Journal for Equity in Health,* 2014, 13(1), 1-12.

[12] E. Hadziabdic, "The use of Interpreter in Healthcare: Perspectives of Individuals, Healthcare Staff and Families," Linnaeus University Press, 2011.

[13] E. C. Khoong, E. Steinbrook, C. Brown, and A. Fernandez, "Assessing the use of Google Translate for Spanish and Chinese Translations of Emergency Department Discharge Instructions," *JAMA Internal Medicine,* 2019, 179(4), 580-582.

[14] G. Flores, "The Impact of Medical Interpreter Services on the Quality of Healthcare: a Systematic Review," *Medical Care Research and Review,* 2005, 62(3), 255-299.

[15] S. Giordano, "Overview of the Advantages and Disadvantages of Professional and Child Interpreters for Limited English Proficiency Patients in General Health Care Situations," *Journal of Radiology Nursing,* 2007, 26(4), 126-131.

[16] J. A. Rodriguez, A. Fossa, R. Mishuris, and B. Herrick, "Bridging the Language Gap in Patient Portals: An Evaluation of Google Translate," *Journal of General Internal Medicine,* 2020, 1-3.

[17] M. Alhamami, "Language barriers in multilingual Saudi hospitals: Causes, consequences, and solutions," *International Journal of Multilingualism,"* 2020, 1-13.

[18] J. F. Ha and L. Nancy, "Doctor-patient communication: a review," *Ochsner Journal,* 2010, 10(1), 38-43.

[19] L. ML. Ong, J. CJM. De Haes, A. M. Hoos, and F. B. Lammes, "Doctor-patient communication: a review of the literature," *Social Science and Medicine,* 1995, 40(7), 903-918.

[20] P. A. Ali and J. Stacy, "Speaking my patient's language: bilingual nurses' perspective about provision of language concordant care to patients with limited English proficiency," *Journal of advanced nursing,* 2017, 73(2), 421-432.

[21] P. Crawford, B. Brown, and K. Harvey, "Corpus linguistics and evidence-based health communication," *The Routledge handbook of language and health communication,* 2014, 75-90.

[22] S. Adolphs, ”Applying corpus linguistics in a health care context,” *Journal of applied linguistics,* 2004, 1(1).

[23] D. Knight, ”A multi-modal corpus approach to the analysis of backchanneling behaviour,” 2009, Dissertation, University of Nottingham.

[24] Defined.AI. The Challenge of Building Corpus for NLP, 2020, Librarieshttps://www.defined.ai/blog/the-challenge-of-building-corpus-for-nlp-libraries/. Accessed on May 22, 2022

[25] A. Elnagar, ”Systematic literature review of dialectal Arabic: identification and detection,” *IEEE Access,* 2021, 9, 31010-31042.

[26] K. Almeman and M. Lee, “Automatic building of Arabic multi dialect text corpora by bootstrapping dialect words,” in Proceedings of 1st International Conference on Communications, Signal Processing, and their Applications. (ICCSPA), Feb. 2013, 1–6.

[27] R. Boujelbane, M. E. Khemekhem, S. BenAyed, and L. H. Belguith, “Building bilingual lexicon to create Dialect Tunisian corpora and adapt language model,” in Proceedings of 2nd Workshop Hybrid Approaches Translation, 2013, 88–93.

[28] R. Al-Sabbagh and R. Girju, “YADAC: Yet another dialectal Arabic corpus,” in Proceedings of LREC, 2012, 2882–2889.

[29] I. Guellil, ”Arabic natural language processing: An overview,” *Journal of King Saud University-Computer and Information Sciences,* 2021, 33(5), 497-507.

[30] T. Wael, ”Intelligent Arabic-Based Healthcare Assistant,” in Proceedings of 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES), October 2021.

[31] H. Bouamor, N. Habash, M. Salameh, W. Zaghouani, O. Rambow, D. Abdulrahim, O. Obeid, S. Khalifa, F. Eryani, and A. Erdmann, ”The Madar Arabic Dialect Corpus and Lexicon,” in *Proc. Eleventh International Conference on Language Resources and Evaluation,* Myazaki, Japan, 2018.

[32] M. Embarki and M. Ennaji, *Modern Trends in Arabic Dialectology,* Red Sea Press, 2011.

[33] A. Farghaly and K. Shaalan, ”Arabic Natural Language Processing: Challenges and Solutions,” *ACM Transactions on Asian Language Information Processing,* 2009, 8(4), 1-22.

[34] A. Torfi, R.A. Shirvani, Y. Keneshloo, N. Tavvaf, and E. Fox, ”A Natural Language Processing Advancements by Deep Learning: A Survey,” arXiv preprint arXiv:2003.01200, 2020.

[35] M. Al-Ayyoub, A. Nuseir, K. Alsmearat, Y. Jararweh, and B. Gupta, ”Deep Learning for Arabic NLP: A Survey,” *Journal of Computational Science,* 2018, 26, 522-531.

[36] M. A. Omari and M. Al-Hajj, ”Classifiers for Arabic NLP: Survey,” *International Journal of Computational Complexity and Intelligent Algorithms,* 2020, 1(3), 231-258.

[37] Md. S. Islam, Md. M. Hasan, X. Wang, Xiaoyi, and H. D. Germack, ”A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining,” *Healthcare,* 2018, 6(2), 54.

[38] R. Ooms and M. Spruit, ”Self-Service Data Science in Healthcare with Automated Machine Learning,” *Applied Sciences,* 2020, 10(9), 2992.

[39] A. Y. Samarah, ”Politeness in Arabic Culture,” *Theory and Practice in Language Studies,* 2015, 5(10), 2005-2016.

[40] E. Hadziabdic and K. Hjelm, ”Working with Interpreters: Practical Advice for use of an Interpreter in Healthcare,” *International Journal of Evidence-Based Healthcare,* 2013, 11(1), 69-76.

[41] M. Alhamami, ”Switching of Language Varieties in Saudi Multilingual Hospitals: Insiders’ Experiences,” *Journal of Multilingual and Multicultural Development,* 2020, 41(2), 175-189.

[42] P. Ram, ”Management of Healthcare in the Gulf Cooperation Council (GCC) Countries with Special Reference to Saudi Arabia,” *International Journal of Academic Research in Business and Social Sciences,* 2014, 4(12), 24.

[43] T. Khoja, S. Rawaf, W. Qidwai, D. Rawaf, K. Nanji, and A. Hamad, ”Healthcare in Gulf Cooperation Council Countries: a Review of Challenges and Opportunities,” *Cureus,* 2017, 9(8).

[44] L. Abualigah, H. E. Alfar, M. Shehab, A. Hussein, and M. A. Alhareth, ”Sentiment Analysis in Healthcare: a Brief Review,” *Recent Advances in NLP: The Case of Arabic Language,* 2020, 129-141.

[45] D.M.E.D.M Hussein, ”A Survey on Sentiment Analysis Challenges,” *Journal of King Saud University - Engineering Sciences,* 2018, 30(4), 330-338.

[46] K. Denecke and Y. Deng, ”Sentiment Analysis in Medical Settings: New Opportunities and Challenges,” *Artificial Intelligence in Medicine,* 2015, 64(1), 17-27.

[47] C. L. Ventola, ”Social Media and Health Care Professionals: Benefits, Risks, and Best Practices,” *Pharmacy and Therapeutics*, 2014, 39(7), 491-499.

[48] A. Alnawas, ”The Corpus Based Approach to Sentiment Analysis in Modern Standard Arabic and Arabic Dialects: A Literature Review,” *Politeknik Dergisi*, 2018, 21(2), 461-470.

[49] G. Mendels, “How to Apply Machine Learning and Deep Learning Methods to Audio Analysis,” Medium, Towards Data Science, November 18, 2019, https://towardsdatascience.com/how-to-apply-machine-learning-and-deep-learning-methods-to-audio-analysis-615e286fcbbc. Accessed February 8, 2021.

[50] T. Cooper, “How Shazam Works,” Medium, 29 January, 2018, medium.com/@treycoopermusic/how-shazam-works-d97135fb4582. Accessed February 8, 2021.

[51] J.A. Reyes-Ortiz, B.A. Gonzalez-Beltran, and L. Gallardo-Lopez, ”Clinical Decision Support Systems: a Survey of NLP-Based Approaches from Unstructured Data,” in *26th International Workshop on Database and Expert Systems Applications (DEXA),* Valencia, Spain, September, 2015.

[52] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, ”Secure and Robust Machine Learning for Healthcare: A Survey,” in arXiv, 2020.

[53] M. Nunez, ”Medical Translation and Artificial Intelligence,” *SimulTrans*, 2020. https://www.simultrans.com/blog/medical-translation-artificial-intelligence. Accessed February 24, 2021.

[54] J. TeCho, ”A Corpus-Based Approach for Keyword Identification using Supervised Learning Techniques,” in *Proc. 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology,* Krabi, Thailand, 2008.

[55] V. Bansal, G. Pahwa, and N. Kannan, ”Cough Classification for COVID-19 Based on Audio Mfcc Features using Convolutional Neural Networks,” in *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON),* India, 2020.

[56] S. Arjun, A. Biswas, A. Gandhi, S. Patil, and O. Deshmukh, ”LIVELINET: A Multimodal Deep Recurrent Neural Network to Predict Liveliness in Educational Videos,” *International Educational Data Mining Society,* 2016.

[57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ”Imagenet Classification with Deep Convolutional Neural Networks,” *Advances in neural information processing systems*, 2012, 25, 1097-1105.

[58] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[59] S. J. Pan and Q. Yang, ”A Survey on Transfer Learning,” *IEEE Transactions on Knowledge & Data Engineering,* 2009, 22(10), 1345-1359.

[60] P. T. Chen, C. L. Lin, and W. N. Wu, ”Big Data Management in Healthcare: Adoption Challenges and Implications,” *International Journal of Information Management,* 2020, 53, 102078.

[61] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, ”Speech Recognition Using Deep Neural Networks: A Systematic Review,” *IEEE Access,* 2019, 7, 19143-19165, doi: 10.1109/ACCESS.2019

[62] N. Haeri, ”Form and Ideology: Arabic Sociolinguistics and Beyond,” *Annual Review of Anthropology,* 200, 29, 61-87.

[63] V. Sergon. Expatica. Introduction to Arabic: the Language of the UAE, March 11, 2021, https://www.expatica.com/ae/education/language-learning/introduction-to-arabic-the-language-of-the-united-arab-emirates-71422. Accessed February 11, 2021.

[64] F.V. Rosse, M.D. Bruijne, J. Suurmond, M. EssinkBot, and C. Wagner, ”Language Barriers and Patient Safety Risks in Hospital Care. A mixed Methods Study,” *International Journal of Nursing Studies,* 2016, 54, 45–53.

[65] S. Fares, F. B Irfan, R. F. Corder, M. A. AlMarzouqi, A. H. AlZaabi, M. M. Idrees, and M. Abbo, ”Emergency Medicine in the United Arab Emirates,” *International Journal of Emergency Medicine*, 2014, 7(4).

[66] S. Patil and P. D. Sumant, ”Use of Google Translate in Medical Communication: Evaluation of Accuracy,” *The BMJ,* 2016, 349(7392).

[67] V. Boddapatia, A. Petefb, and J.L.L. Rasmussonb, ”Classifying environmental sounds using image recognition networks,” *Procedia Computer Science,* 2017, 112, 2048–2056.

[68] S. Seo and S.B. Cho,”Offensive Sentence Classification using Character-Level CNN and Transfer Learning with Fake Sentences,” in *International Conference on Neural Information Processing,* Guangzhou, China, November, 2017.

[69] D. Wang and T.F. Zheng, ”Transfer Learning for Speech and Language Processing,” in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA),* Hong Kong Polytechnic University, Hong Kong, December, 2015.

[70] B. Sertolli, R. Zhao, B.W. Schuller, and N. Cummins, ”Representation transfer learning from deep end-to-end speech recognition networks for the classification of health states from speech,” *Computer Speech & Language*, 2021, 68, 101204.

[71] J.C. Vásquez-Correa, C.D. Rios-Urrego, T. Arias-Vergara, M. Schuster, J. Rusz, E. Nöth, and J.R. Orozco-Arroyave, ”Transfer Learning Helps to Improve the Accuracy to Classify Patients with Different Speech Disorders in Different Languages,” *Pattern Recognition Letters*, In press, 2021.

[72] Y. Chen, B. Gao, L. Jiang, K. Yin, J. Gu and W.L. Woo, and Wai Lok, ”Transfer learning for wearable long-term social speech evaluations,” *IEEE Access*, 2018, 6, 61305-61316.

[73] K. Feng and T. Chaspari, ”Low-resource language identification from speech using transfer learning,” in *29th International Workshop on Machine Learning for Signal Processing (MLSP),* Pittsburg, PA, US, October, 2019.

[74] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, and L. Fei-Fei, ”Imagenet: A large-scale hierarchical image database,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Florida, US, June, 2009.

[75] A. Senn, ”How did participants experience volunteering for a translation crowdsourcing project?,” PhD dissertation, University of Geneva, 2021.

[76] X. Chen, L. Zhengxiong, S. Srirangaraj, and X. Wenyao, ”Exploring racial and gender disparities in voice biometrics,” *Scientific Reports,* 2022, 12(1), 1-12.

[77] S. Taylor, C. Dromey, S.L. Nissen, K. Tanner, D. Eggett, and K. Corbin-Lewis, ”Age-related changes in speech and voice: spectral and cepstral measures,” *Journal of Speech, Language, and Hearing Research,*, 2020, 63(3), 647-660.

[78] E. West, ”Ethics and integrity in nursing research,” *Handbook of Research Ethics and Scientific Integrity,* 2020, 1051-1069.

Design of Higher-Dimensional Hyperchaotic System based on Combined Control and its Encryption Application

Kun Zhao, Jianbin He*
School of Mathematics and Statistics
Minnan Normal University, Zhangzhou 363000, China

Abstract—**According to the anti-control principle of chaos, a combined control method is proposed based on a class of asymptotically stable linear systems with multiple controllers. A higher-dimensional hyperchaotic system is investigated by the Lyapunov exponents method and equilibrium points analysis, and it exists the largest number of positive Lyapunov exponents. The chaotic pseudo-random sequences of the higher-dimensional hyperchaotic system can pass all NIST tests after preprocessing, and behave better chaotic characteristics. Meanwhile, a new encryption algorithm of image information with position scrambling, sequential diffusion and reverse diffusion is designed based on the chaotic pseudo-random sequences. The experiments of image information are given to verify the effectiveness and feasibility of the encryption algorithm. Finally, the security analyses are also discussed by the key sensitivity, differential attack and statistical analysis. It is shown that the encryption algorithm has large enough key space and can be applied to secure communication.**

Keywords—***Hyperchaotic system; positive lyapunov exponent; chaotic pseudo-random sequence; image encryption***

I. Introduction

With the rapid development of computer network technology and intelligent equipment, the digital information may be stolen or even destroyed by the attacker when it is transmitted through the public network, such as the personal privacy information, images and videos. In particular, some information is used in military, medical, political and other important fields, so it is very necessary to protect the integrity and confidentiality of the information transmission process. Information hiding and information encryption are two important information protection technologies. Information hiding is to hide information in another information carrier and transmit it through public channel, and it includes information hiding algorithm, digital watermarking, hidden channel technology and anonymous communication technology, etc. The information encryption is to design encryption algorithms to improve the security and efficiency by the characteristics of digital information. Usually, the image encryption includes uncompressed and compressed image encryption [1].

The chaos-based image encryption is one of widely used security method. It can not only prevent the loss of image information, but also convert original image into unrecognized encrypted image. The chaos-based image encryption generally includes two important steps: scrambling and diffusion encryption. The scrambling method can scramble the position of the plaintext image without changing the pixel value of the image, and it reduces the correlation between adjacent pixels of the image. The diffusion method changes the pixel value of the image through XOR operation, which makes the distribution of the encrypted image information more uniform and random. Therefore, the combination of scrambling and diffusion encryption is very effective for the image encryption.

Since Lorenz found the chaos from the mathematical model of meteorology, chaos theory has attracted extensive attention of scientists. Chaotic system is usually generated from a nonlinear dynamic system, and the characteristics of initial condition sensitivity, non-periodicity, long-term unpredictability and pseudo-randomness are very suitable for image encryption and other information encryption [2]–[7]. Moreover, some of the chaos-based encryption algorithms are analysed and may be not resist the chosen-plaintext attacks [8]. In [9], the security loopholes of an image encryption algorithm based on random walk and hyperchaotic systems are found, and the attack method is proposed to successfully break the encryption scheme. Therefore, the security of encryption algorithm based on chaotic system is one of the most important factors for information secure communication, and more secure chaos-based encryption algorithm need to be analysed and proposed. Compared with the lower-dimensional chaotic system, the higher-dimensional hyperchaotic system has more positive Lyapunov exponents, and the pseudo-random sequences generated by iteration are more complex chaotic characteristics. The encryption algorithm based on the higher-dimensional hyperchaotic system can be used for information secure communication [10]. The positive Lyapunov exponent is one of useful methods to show whether the nonlinear dynamic system exists chaos or not. Generally, the chaotic system has one positive Lyapunov exponent, while the hyperchaotic system has two or more positive Lyapunov exponents [11]. The number and size of positive Lyapunov exponents can reflect the chaotic characteristics of the system, and the hyperchaotic system with multiple positive Lyapunov exponents has more complex dynamic characteristics.

In recent years, the research on higher-dimensional hyperchaotic systems with multiple positive Lyapunov exponents has attracted much attention [12]–[14]. In [15], an effective image encryption algorithm of confusion and diffusion encryption is proposed based on chaotic system, and it is very sensitive to the initial variables. A new chaos-based image encryption algorithm is investigated in [16], and the security test shows

*Corresponding authors.

that the algorithm has good security performance and can resist a variety of special attacks. An image encryption algorithm based on chaotic system and DNA sequence operation is proposed, which not only has good encryption effect, but also can resist various typical attacks [17]. A S-box encryption algorithm based on chaotic system is designed for secure and fast image encryption, and NIST tests are used to verify the randomness of the sequences [18]. A color image encryption scheme is proposed based on non-uniform cellular automata and hyperchaos, the security analysis shows that the scheme has a very large key space and can resist various attacks [19]. A new image encryption algorithm is proposed by the confusion and diffusion based on chaos and SHA-256, and it can resist the chosen-plaintext attack and overcome the difficulty of key management in the "one-time password" encryption scheme [20]. A modified logistic chaotic map are created to designed encryption technique with better security and efficiency [21]. An image encryption algorithm is designed by combining fractional Fourier transform, DNA sequence operation and chaos theory, and the algorithm has good encryption effect, large key space and high key sensitivity [22]. A technique for encrypting RGB image components by using nonlinear chaotic function and DNA sequence is presented in [23]. In [24], the theoretical security of a medical privacy protection scheme based on DNA encoding and chaotic maps is reanalyzed, and the scheme is rigorously proven to be insecure against the chosen-plaintext attack. Based on a combination of multidimensional chaotic systems, an cryptosystem for the color image encryption is described in [25], and the level of security and the computational complexity is improved. By using higher-dimensional chaotic maps and some conventional cryptographic techniques, a class of chaotic cryptosystems is designed to enhance the security of cryptosystems [26].

The research of higher-dimensional hyperchaotic systems is one of hot topic. Some criteria and methods for constructing higher-dimensional hyperchaotic systems are proposed [27], [28]. In this paper, a higher-dimensional hyperchaotic system is investigated by the combination of multiple controllers, and a new 11-dimensional hyperchaotic system with nine positive Lyapunov exponents is designed. The main contributions of this paper are as follows: (1) Through the combination of multiple controllers, a class of higher-dimensional hyperchaotic systems with the largest number of positive Lyapunov exponents is studied; (2) Based on the chaotic sequences generated by the iteration of higher-dimensional hyperchaotic system, an encryption algorithm is proposed by combining the position scrambling, sequential diffusion and inverse diffusion. (3) The feasibility and security of the new encryption algorithm based on 11-dimensional hyperchaotic system are verified through the simulation experiments.

The rest of this paper is organized as follows. Section II is the design method of hyperchaotic system. Section III is the design of encryption algorithm. The security analysis of the encryption algorithm is given in Section IV. Section V concludes the paper.

II. Construction of Higher-Dimensional Hyperchaotic Systems

A. Chaotic Anti-Control System with Combined Controllers

According to the anti-control method of higher-dimensional hyperchaotic systems, a nominal asymptotically stable linear dynamical system is given by [29]

$$\dot{X} = PAP^{-1}X \tag{1}$$

where $X = (x_1, x_2, \cdots, x_n)$, the matrix A and the similarity transformation matrix P are given as follows:

$$\left\{ \begin{array}{l} A = \begin{pmatrix} A_1 & 0 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_{m-1} & 0 \\ 0 & 0 & \cdots & 0 & A_m \end{pmatrix}_{n\times n} \\ \qquad \text{if } n \text{ is even}, m = \dfrac{n}{2} \\ A = \begin{pmatrix} A_1 & 0 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_m & 0 \\ -1 & -1 & \cdots & -1 & \tau \end{pmatrix}_{n\times n} \\ \qquad \text{if } n \text{ is odd}, m = \dfrac{n-1}{2} \end{array} \right.$$

$$P = \begin{pmatrix} 0 & 1 & \cdots & 1 & 1 \\ 1 & 0 & \cdots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}_{n\times n}$$

and $A_i = \begin{pmatrix} \lambda_i & \psi_{i1} \\ \psi_{i2} & \lambda_i \end{pmatrix}$

is a block matrix, where $\psi_{i1} \times \psi_{i2} < 0, \tau < 0$.

Next, a uniformly bounded controller $f(\sigma X, \varepsilon)$ and control matrix C are designed for the system (1), such that

$$\dot{X} = PAP^{-1}X + Cf(\sigma X, \varepsilon) \tag{2}$$

The combination of controllers $f(\sigma X, \varepsilon)$ and the control matrix C are given by

$$f(\sigma X, \varepsilon) = \begin{pmatrix} \varepsilon_1 \sin(\sigma_1 x_1 + \varphi_1) + \varepsilon_2 \cos(\sigma_2 x_1 + \varphi_2) \\ \varepsilon_1 \sin(\sigma_1 x_2 + \varphi_1) + \varepsilon_2 \cos(\sigma_2 x_2 + \varphi_2) \\ \vdots \\ \varepsilon_1 \sin(\sigma_1 x_{n-1} + \varphi_1) + \varepsilon_2 \cos(\sigma_2 x_{n-1} + \varphi_2) \\ \varepsilon_1 \sin(\sigma_1 x_n + \varphi_1) + \varepsilon_2 \cos(\sigma_2 x_n + \varphi_2) \end{pmatrix}$$

$$C = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & 1_{(i,j)} & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{n\times n}$$

where $\varepsilon_1, \sigma_1, \varphi_1, \varepsilon_2, \sigma_2, \varphi_2$ are controller parameters, $1_{(i,j)}$ denotes that the element in row i and column j is equal to 1, i.e., the controller is in the state of working.

When the dimension $n = 11$, the matrices A_i $(i = 1, 2, 3, 4, 5)$ are given by

$$\begin{cases} A_1 = \begin{pmatrix} -0.01 & 1.00 \\ -6.00 & -0.01 \end{pmatrix}, A_2 = \begin{pmatrix} -0.01 & 18.00 \\ -2 & -0.01 \end{pmatrix} \\ A_3 = \begin{pmatrix} -0.01 & 15.00 \\ -1.00 & -0.01 \end{pmatrix}, A_4 = \begin{pmatrix} -0.01 & 22.00 \\ -2.50 & -0.01 \end{pmatrix} \\ A_5 = \begin{pmatrix} -0.01 & 3.00 \\ -20.00 & -0.01 \end{pmatrix} \end{cases}$$

and $\tau = -0.01, \varepsilon_1 = 76, \sigma_1 = 8, \varphi_1 = 6, \varepsilon_2 = 68, \sigma_2 = 5, \varphi_2 = 4$, the control position $(i, j) = (11, 10)$, therefore, the controlled system is given as follows:

$$\dot{X} = PAP^{-1}X + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ f(x_{10}) \end{bmatrix}_{11\times 1} \tag{3}$$

where the combined controller

$$f(x_{10}) = 76\sin(8x_{10} + 6) + 68\cos(5x_{10} + 4)$$

and it is shown in Fig. 1.

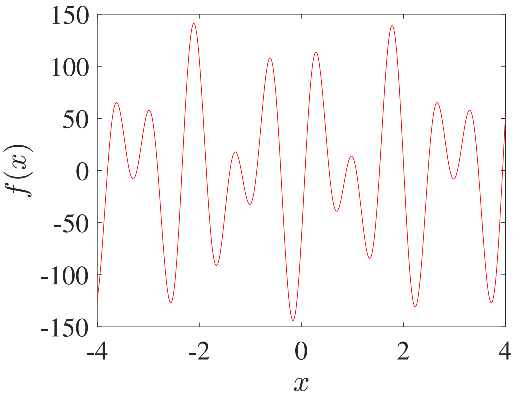


Fig. 1. The Function of Combined Controller $f(x_{10})$ (x Denotes x_{10})

B. Chaotic Attractors and Lyapunov Exponent

By the calculation on the software of Matlab R2021a, the Lyapunov exponents of system (3) are given by

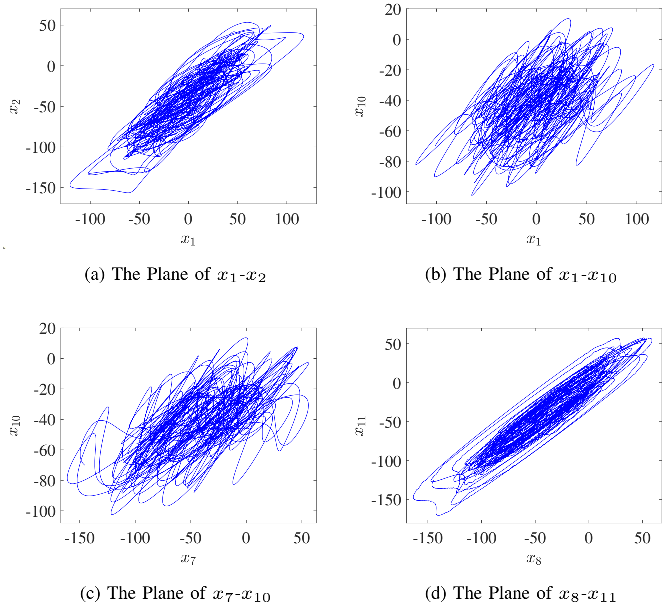
$$\begin{cases} \mathrm{LE}_1 = 4.37, \mathrm{LE}_2 = 0.49, \mathrm{LE}_3 = 0.41 \\ \mathrm{LE}_4 = 0.35, \mathrm{LE}_5 = 0.29, \mathrm{LE}_6 = 0.26 \\ \mathrm{LE}_7 = 0.22, \mathrm{LE}_8 = 0.16, \mathrm{LE}_9 = 0.02 \\ \mathrm{LE}_{10} = 0.00, \mathrm{LE}_{11} = -6.68 \end{cases}$$

Obviously, the 11-dimensional hyperchaotic system has 9 positive Lyapunov exponents, so it has strong chaotic characteristics.

Furthermore, the initial values

$$X(0) = (0.2, 0.1, 0.3, 0.1, 0.2, 0.1, 0.5, 0.6, 0.7, 0.4, 0.2)$$

then the phase diagrams of chaotic attractor are shown in Fig. 2.



(a) The Plane of x_1-x_2

(b) The Plane of x_1-x_{10}

(c) The Plane of x_7-x_{10}

(d) The Plane of x_8-x_{11}

Fig. 2. Attractor of the Controlled Hyperchaotic System (3)

C. Equilibrium Point Analysis of the Controlled System

Obviously, the only one equilibrium point of the system (1) is

$$X_e = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$

and the corresponding eigenvalues of Jacobi matrix at the equilibrium point X_e are given by

$$\begin{cases} \lambda_{1,2} = -0.01 \pm 7.7460\mathrm{i} \\ \lambda_{3,4} = -0.01 \pm 7.4162\mathrm{i} \\ \lambda_{5,6} = -0.01 \pm 6.0000\mathrm{i} \\ \lambda_7 = -0.01 \\ \lambda_{8,9} = -0.01 \pm 3.8730\mathrm{i} \\ \lambda_{10,11} = -0.01 \pm 2.4495\mathrm{i} \end{cases}$$

Since all eigenvalues are negative, so the system (1) is asymptotically stable.

However, the controlled system (3) has multiple equilibrium points, and the equilibrium points of the controlled system (3) can be obtained by Eq. (4).

Therefore, the corresponding solutions can be obtained by the Cramer rule [11]:

$$\begin{aligned} x_i &= \frac{|E_i|}{|D|} \\ &= (-1)^{(11+i)} \frac{-f(x_{10})}{|D|} |M_{11,i}| \\ &\quad (i = 1, 2, \cdots, 11) \end{aligned}$$

where $D = PAP^{-1}$, E_i is the matrix that the ith column of the matrix D is replaced by the controller vector in right-

$$\begin{pmatrix} 2.74 & -0.25 & 22.75 & -19.25 & 5.25 & -12.25 & 3.75 & -15.25 & 4.75 & 1.75 & 8.75 \\ 3.75 & 1.74 & 4.75 & -17.25 & 7.25 & -10.25 & 5.75 & -13.25 & 6.75 & 3.75 & 10.75 \\ 1.45 & 2.45 & 22.44 & -19.55 & 4.95 & -12.55 & 3.45 & -15.55 & 4.45 & 1.45 & 8.45 \\ 2.00 & 0.00 & 23.00 & -19.01 & 3.00 & -12.00 & 4.00 & -15.00 & 5.00 & 2.00 & 9.00 \\ -0.45 & -2.45 & 20.55 & 0.55 & 3.04 & -14.45 & 1.55 & -17.45 & 2.55 & -0.45 & 6.55 \\ 1.85 & -0.15 & 22.85 & -19.15 & 5.35 & -12.16 & 2.85 & -15.15 & 4.85 & 1.85 & 8.85 \\ 0.25 & -1.75 & 21.25 & -20.75 & 3.75 & 1.25 & 2.24 & -16.75 & 3.25 & 0.25 & 7.25 \\ 1.95 & -0.05 & 22.95 & -19.05 & 5.45 & -12.05 & 3.95 & -15.06 & 2.95 & 1.95 & 8.95 \\ -0.05 & -2.05 & 20.95 & -21.05 & 3.45 & -14.05 & 1.95 & 0.95 & 2.94 & -0.05 & 6.95 \\ 2.35 & 0.35 & 23.35 & -18.65 & 5.85 & -11.65 & 4.35 & -14.65 & 5.35 & 2.34 & 3.35 \\ 1.65 & -0.35 & 22.65 & -19.35 & 5.15 & -12.35 & 3.65 & -15.35 & 4.65 & 2.65 & 8.64 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ f(x_{10}) \end{pmatrix} \tag{4}$$

hand side of the Eq. (4), and $M_{11,i}$ is the algebraic cofactor of E_i, i.e., the solutions x_i of the controlled system (3) are given by

$$\begin{cases} x_1 = \dfrac{11339f(x_{10})}{-106923}, & x_2 = \dfrac{10152750f(x_{10})}{106923} \\ x_3 = \dfrac{10156849f(x_{10})}{106923}, & x_4 = \dfrac{10155828f(x_{10})}{106923} \\ x_5 = \dfrac{10160590f(x_{10})}{106923}, & x_6 = \dfrac{10155606f(x_{10})}{106923} \\ x_7 = \dfrac{10167011f(x_{10})}{106923}, & x_8 = \dfrac{10155721f(x_{10})}{106923} \\ x_9 = \dfrac{10161661f(x_{10})}{106923}, & x_{10} = \dfrac{10252559f(x_{10})}{106923} \\ x_{11} = \dfrac{10157933f(x_{10})}{106923} \end{cases} \tag{5}$$

Hence, one has

$$\frac{106923}{10252559}x_{10} = -76\sin(8x_{10}+6) - 68\cos(5x_{10}+4)$$

if one lets

$$y_1 = \frac{106923}{10252559}x_{10}$$

and

$$y_2 = -76\sin(8x_{10}+6) - 68\cos(5x_{10}+4)$$

then the intersection points (x_{10}, y_1) of $y_1 = y_2$ are shown in Fig. 3, and the equilibrium points of the controlled system are given by Eq. (5).

III. Design of Encryption Algorithm

A. Data Preprocessing

An image encryption scheme is designed based on 11-dimensional hyperchaotic system. Firstly, the 4th-order Runge-Kutta method is used to discretize the 11-dimensional hyperchaotic system, where the Runge-Kutta formula is given by

$$\begin{cases} X_{i+1} = X_i + \dfrac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4) \\ K_1 = f(t_i, X_i), K_2 = f(t_i + \dfrac{h}{2}, X_i + \dfrac{h}{2}K_1) \\ K_3 = f(t_i + \dfrac{h}{2}, X_i + \dfrac{h}{2}K_2), K_4 = f(t_i + h, X_i + hK_3) \\ \quad (i = 1, 2, \cdots n, \cdots) \end{cases}$$

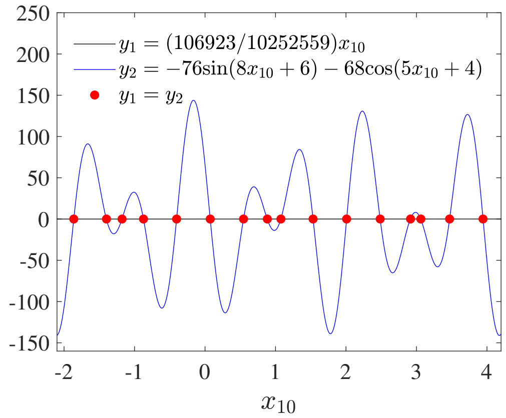


Fig. 3. The Intersection Points of y_1 and y_2 in Equation (4)

The initial values of the hyperchaotic system $X(0)$, and the step $h = 0.001$, then the image encryption algorithm are given as follows:

Step 1: The number of pre-iterations is equal to $\big(2000 + \mathrm{mod}(\mathrm{sum}, \sigma_1 \times \varepsilon_1 \times \varphi_1)\big)$ and it is used to counteract the transient effect of chaotic iteration, where sum is sum of pixel values of original image, mod is the modular function, and $\sigma_1, \varepsilon_1, \varphi_1$ are controller parameters. The pseudo-random sequences generated by the iteration of 11-dimensional hyperchaotic system are $X = (X_1, X_2, \cdots, X_{11})$.

Step 2: The operations of rounding, modulo and shifting are used to generate the pseudo-random sequences $Z =$

$(Z_1, Z_2, \cdots, Z_{11})$, i.e.,

$$\begin{cases} Z_1 = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_1, 1) \times 10^{12}, 1) \times 10^9) \\ Z_2 = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_2, 1) \times 10^{14}, 1) \times 10^9) \\ Z_3 = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_3, 1) \times 10^{13}, 1) \times 10^9) \\ Z_4 = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_4, 1) \times 10^{13}, 1) \times 10^9) \\ Z_5 = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_5, 1) \times 10^{12}, 1) \times 10^9) \\ Z_6 = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_6, 1) \times 10^{13}, 1) \times 10^9) \\ Z_7 = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_7, 1) \times 10^{12}, 1) \times 10^9) \\ Z_8 = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_8, 1) \times 10^{14}, 1) \times 10^9) \\ Z_9 = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_9, 1) \times 10^{11}, 1) \times 10^9) \\ Z_{10} = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_{10}, 1) \times 10^{13}, 1) \times 10^9) \\ Z_{11} = \mathrm{fix}(\mathrm{mod}(\mathrm{mod}(X_{11}, 1) \times 10^{15}, 1) \times 10^9) \end{cases}$$

where the function fix represents the rounding operation, and the pseudo-random sequences Z can pass most tests of NIST.

Step 3: In order to ensure that the encryption algorithm has better encryption effect, the pseudo-random sequences Z are further obtained by

$$\begin{cases} W_1 = \mathrm{mod}((Z_1 - Z_2 + Z_3), MN) + 1 \\ W_2 = \mathrm{mod}((Z_4 + Z_5), 256) \\ W_3 = \mathrm{mod}((Z_6 + Z_7 + 1), 256) \\ W_4 = \mathrm{mod}((Z_8 + Z_9 + 1), 256) \\ W_5 = \mathrm{mod}((Z_{10} + Z_{11}), 256) \end{cases}$$

Then the new pseudo-random sequences $W = (W_1, W_2, \cdots, W_5)$ can pass the NIST test, and they are given in Section V.

B. Encryption Algorithms

The encryption algorithms include scrambling encryption, sequential diffusion encryption and reverse diffusion encryption. The image information is chosen as an example, and the flow chart of information encryption is shown in Fig. 4.

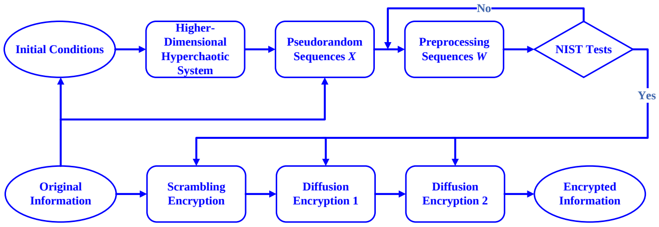


Fig. 4. The Flow Chart of Information Encryption

a) Position Scrambling Encryption

The original image P_1 is in the size of $M \times N$, and the pixel position of P_1 is scrambled based on the pseudo-random sequence W_1, then the scrambled image P_2 is obtained by

$$P_2(i) = P_1(W_1(i)), \ (i = 1, 2, \cdots, MN)$$

b) Sequential Diffusion Encryption

The scrambled image P_2 is encrypted by using pseudo-random sequences W_3 and W_4, and the steps of encryption are as follows:

Step 1: The first pixel value $P_2(1)$ of the scrambled image is encrypted by the first value $W_3(1)$ of the random sequence via the XOR operation, i.e.,

$$P_3(1) = P_2(1) \oplus W_3(1)$$

Step 2: Add the pseudo-random sequence $W_2(i)$ to the pixel values of scrambled image $P_2(i)$, and subtract the integer part of $P_2(i-1)/\varphi$, then the encrypted information $P_3'(i)$ is obtained by the modulus of 256, i.e.,

$$P_3'(i) = \mathrm{mod}((P_2(i) + W_2(i) - \mathrm{fix}(P_2(i-1)/\varphi)), 256) \\ (i = 2, 3, \cdots, MN)$$

Step 3: The sequence $P_3^{'}$ is encrypted with the random sequence W_3 by the XOR operation, and the encrypted image is given by

$$P_3(i) = P_3'(i) \oplus W_3(i), \quad (i = 2, 3, \cdots, MN)$$

c) Reverse Diffusion Encryption

Use the random sequence W_4 and W_5 to perform reverse diffusion encryption on the sequential diffusion encrypted image P_3, and the encryption steps are given as follows:

Step 1: The $P_3(MN)$ is encrypted by the pseudo-random sequence $W_5(MN)$, i.e.,

$$P_4(MN) = P_3(MN) \oplus W_5(MN).$$

Step 2: Through subtraction, multiplication and modulo operations, the pixel values of P_3 are encrypted from the pseudo-random sequence W_4, i.e.,

$$P_4'(i) = \mathrm{mod}(W_4(i) - P_3(i) + P_3(i+1), 256) \\ (i = MN-1, MN-2, \cdots, 1)$$

Step 3: Similarly, the P_4' is encrypted by the pseudo-random sequence W_5 by the XOR operation, one has

$$P_4(i) = P_4'(i) \oplus W_5(i), (i = MN-1, MN-2, \cdots, 1)$$

C. Decryption Process

The decryption is the inverse operation of the encryption, and it is given in follows:

Step 1: The $P_4(MN)$ and $W_5(MN)$ is used to obtain the value $P_3(MN)$ of sequential diffusion encryption, i.e.,

$$P_3(MN) = P_4(MN) \oplus W_5(MN)$$

Step 2: By addition, subtraction, multiplication and modulo operations, the P_3 is decrypted by the pseudo-random sequences W_4 and W_5, and it is given by

$$P_3(i) = \mathrm{mod}(W_4(i) + P_4(i+1) - (P_4(i) \oplus W_5(i)), 256) \\ (i = MN-1, MN-2, \cdots, 1)$$

Step 3: Similarly, the $P_2\left(1\right)$ of scrambled encrypted image is obtained by the XOR operation of $P_3\left(1\right)$ and $W_3\left(1\right)$, i.e.,

$$P_2(1) = P_3(1) \oplus W_3(1)$$

Step 4: By the XOR, addition, division, subtraction and modulo operation, the scrambled encrypted image P_2 is decrypted by the pseudo-random sequences W_2 and W_3, and it is given by

$$\begin{aligned} P_2(i) = \operatorname{mod}((&P_3(i) \oplus W_3(i) + fix(P_3(i-1)/\varphi) \\ &- W_2(i)), 256), \quad (i = 2, 3, \cdots, MN) \end{aligned}$$

Step 5: The pixel value $P_2(i)$ of the scrambled image is exchanged with the pseudo-random sequence $W_1(i)$, and then one can get the original imageP_1, i.e.,

$$P_1(i) = P_2(W_1(i)), \quad (i = MN, MN-1, \cdots, 1)$$

Therefore, the decryption process is completed, and the receivers can get the recovered information from the ciphertext.

IV. Experimental Results

Based on the 11-dimensional hyperchaotic system in Eq. (3), the numerical simulation results are given by the proposed encryption and decryption algorithm in Section III. The encryption algorithm is tested by the Lena image in Fig. 5 (a) with the size of 512×512 and the Cameraman image in Fig. 5 (e) with the size of 256×256 based on the Matlab software, and the sum of pixel values of Lena image and Cameraman image are 32515895 and 7780728, respectively. By the initial values

$$X(0) = (0.2, 0.1, 0.3, 0.1, 0.2, 0.1, 0.5, 0.6, 0.7, 0.4, 0.2)$$

and other parameters of controlled system in Eq. (3), the results of encryption and decryption are shown in Fig. 5. The encrypted images Fig. 5 (b) and (f) are chaotic and disordered, and the original information can not be distinguished, so the encryption algorithm is effective.

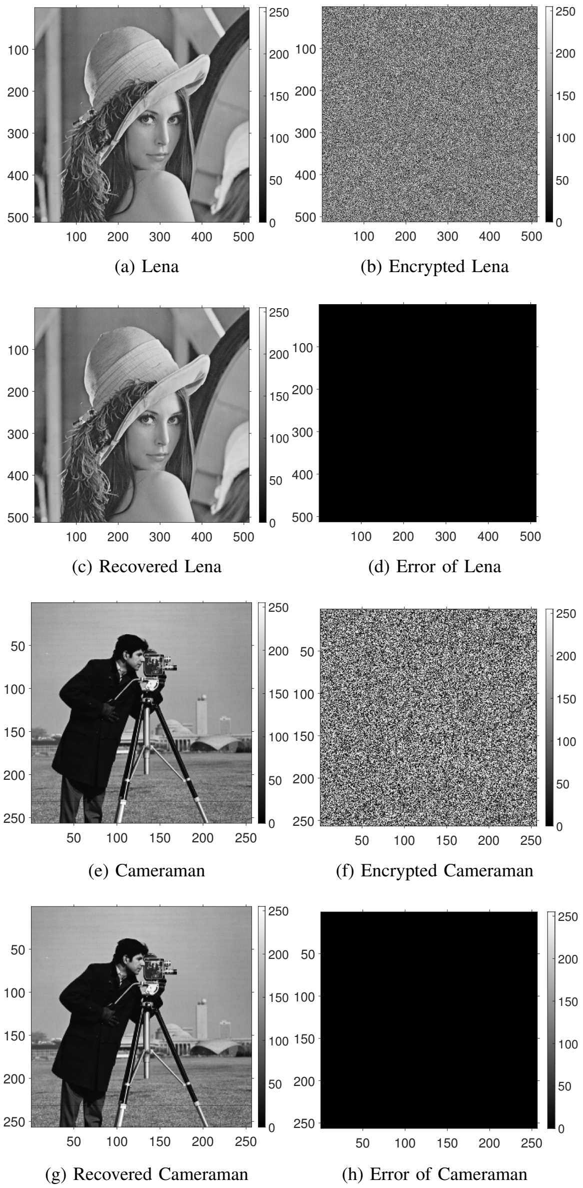
In the encryption algorithm, the sum of image pixel values is used to obtain the key of the encryption. Obviously, the sum of pixel values of different images is different, so the different images will generate different keys to the encryption, i.e., the cryptosystem has the effect of “one-time-pad”. Hence, one cannot get any information of the plaintext image from the encrypted image, and the encryption algorithm is effective for secure communication.

Meanwhile, the error images in Fig. 5 (d) and (h) show that the errors between the recovered image and the original image are equal to zero, and Fig. 5 (c) and (g) show the original information can be recovered successfully by the decryption algorithm.

The hyperchaotic system is highly sensitive to the initial values $X(0), A, P, \varepsilon_i, \sigma_i$ $(i = 1, 2)$, etc., and the initial values are used as the encryption keys. If one of the keys is wrong, the ciphertext image cannot be successfully recovered, because the different initial values will generate different chaotic sequences. For example, if the value of the controller parameter is changed from $\varepsilon_1 = 76$ to $\varepsilon_1 = 75$, then the Lena image is encrypted by the corresponding pseudo-random sequences W, but the experiments show that the Lena image cannot be recovered successfully. Similarly, if the initial values $X(0)$ are changed to

$$X(0) = (0.1, 0.1, 0.3, 0.1, 0.2, 0.1, 0.5, 0.6, 0.7, 0.4, 0.2)$$

the experimental results show that the Lena image cannot be recovered successfully, so the encryption algorithm is also sensitive to the initial values $X(0)$.



(a) Lena

(b) Encrypted Lena

(c) Recovered Lena

(d) Error of Lena

(e) Cameraman

(f) Encrypted Cameraman

(g) Recovered Cameraman

(h) Error of Cameraman

Fig. 5. Experiments Results of Image Encryption and Decryption

V. Security Analysis

A. Analysis of Key Sensitivity

A good encryption algorithm must be sensitive to the small change of the key, i.e., if there is a small change of the key, then the ciphertext image can not recovered completely. The key space is depended on the sensitivity of matrix A, similar transformation matrix P, controller parameter $\varepsilon_1, \sigma_1, \varphi_1$ and the initial values $X(0)$ of controlled system. As there are 121 elements in matrix A and similar transformation matrix P, one needs to keep other keys unchanged but change only one key with small error, and then the experimental results of decipher images are given in Fig. 6. Fig. 6 (a) shows the decrypted image obtained by using the correct keys, and the ciphertext image can be recovered successfully. Fig. 6 (b)-(d) shows the decrypted image with the small error key, but the ciphertext image cannot be recovered successfully. Through experimental tests, Table I shows the ciphertext can not be decrypted successfully when the errors of key is greater than or equal to the minimum values.

TABLE I. TEST RESULTS OF KEY SENSITIVITY

Error of key	Recovered successfully
$\lvert x_1 - x_1' \rvert \geqslant 10^{-16}$	No
$\lvert x_2 - x_2' \rvert \geqslant 10^{-15}$	No
$\lvert x_3 - x_3' \rvert \geqslant 10^{-16}$	No
$\lvert x_4 - x_4' \rvert \geqslant 10^{-17}$	No
$\lvert x_5 - x_5' \rvert \geqslant 10^{-16}$	No
$\lvert x_6 - x_6' \rvert \geqslant 10^{-16}$	No
$\lvert x_7 - x_7' \rvert \geqslant 10^{-16}$	No
$\lvert x_8 - x_8' \rvert \geqslant 10^{-15}$	No
$\lvert x_9 - x_9' \rvert \geqslant 10^{-16}$	No
$\lvert x_{10} - x_{10}' \rvert \geqslant 10^{-16}$	No
$\lvert x_{11} - x_{11}' \rvert \geqslant 10^{-15}$	No
$\lvert \sigma_1 - \sigma_1' \rvert \geqslant 10^{-15}$	No
$\lvert \varepsilon_1 - \varepsilon_1' \rvert \geqslant 10^{-14}$	No
$\lvert \varphi_1 - \varphi_1' \rvert \geqslant 10^{-15}$	No
$\lvert A(i,j) - A(i,j)' \rvert \geqslant 10^{-15}$, $(i,j = 1,2,\cdots,11)$	No
$\lvert P(i,j) - P(i,j)' \rvert \geqslant 10^{-15}$, $(i,j = 1,2,\cdots,11)$	No

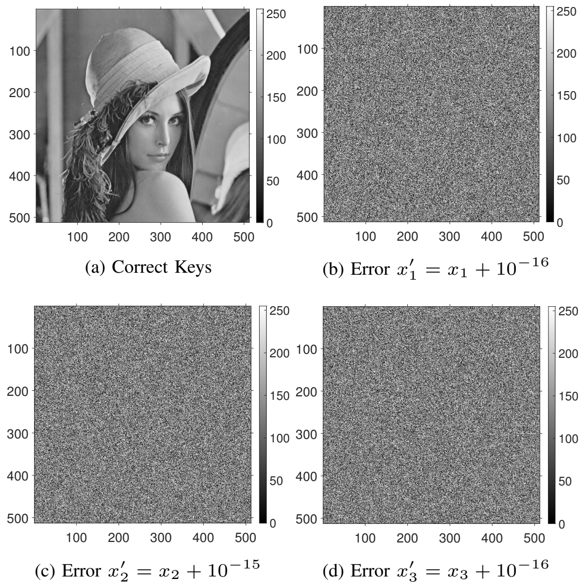
The experimental results show that the image can not be decrypted successfully by using the key with small error. In Table I, it can be estimated that the key space of the encryption algorithm is

$$\begin{aligned} KS &= 10^{14} \times (10^{15})^5 \times (10^{15})^{121} \times (10^{15})^{121} \times (10^{16})^7 \\ &\quad \times 10^{17} \\ &= 10^{3848} \gg 2^{210} \end{aligned}$$

B. Histogram and Chi-Square Test

Histogram describes the distribution of image pixel value. The more uniform the distribution of pixel value, the better the effect of the encryption algorithm. Fig. 7 shows the histograms of the plaintext image and the encrypted image, respectively.

In addition, the Chi-square test is used to illustrate that the cryptosystem has very good confusion characteristics [30].



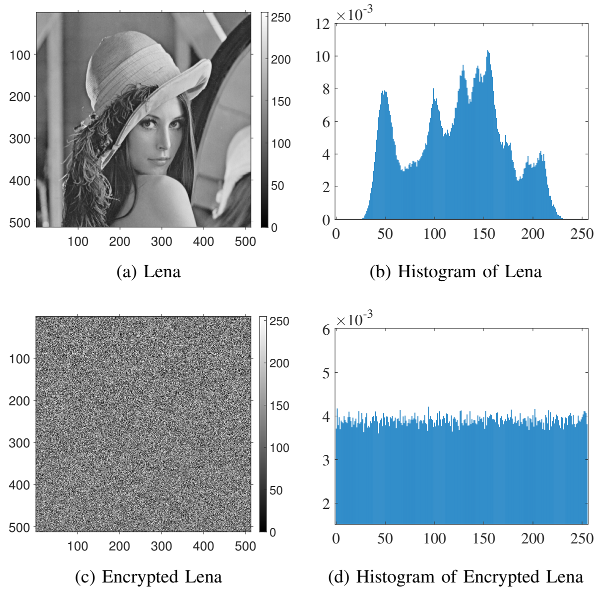
(a) Correct Keys

(b) Error $x_1' = x_1 + 10^{-16}$

(c) Error $x_2' = x_2 + 10^{-15}$

(d) Error $x_3' = x_3 + 10^{-16}$

Fig. 6. Decryption Results of Encrypted Image with Different Keys



(a) Lena

(b) Histogram of Lena

(c) Encrypted Lena

(d) Histogram of Encrypted Lena

Fig. 7. Histogram of Lena and the Corresponding Encrypted Image

The grayscale level of a grayscale image is 256, and then

$$\chi^2 = \sum_{i=0}^{255} \frac{(f_i - g_i)}{g_i}, \quad (i = 0, 1, 2, \cdots, 255)$$

where f_i is the frequency of each 0 to 255 pixel level in the histogram of the encrypted image, g_i is the ideal frequency of uniform distribution, i.e.,

$$g_i = \frac{MN}{256}, \ (i = 0, 1, \cdots, 255)$$

where M and N are the length and width of the image, respectively. If the χ^2 distribution with a degree of freedom of 255 and the significance level is 0.05, then $\chi^2_{0.05}(255) = 293.25$. In Table II, the χ^2 test of original images is significantly greater than 293.25, but the encrypted images are all less than 293.25. Therefore, the distribution of the histogram of the encrypted image is uniform, and it do not disclose any information by the statistical analysis.

TABLE II. χ^2 TESTS

Image	Lena	Cameraman	Barbara
Original Image	158350	110970	144100
Encrypted Image	223.64	238.89	215.80

C. Information Entropy

Information entropy is used to describe the randomness of image information, and it's defined as follows [31]:

$$H = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

where p_i is the probability of the i-th gray value. For gray-scale images, the expected entropy of image information is equal to 8. The entropy of three different images and encrypted images are shown in Table III. The entropy of encrypted images is close to 8, so the encryption algorithm is suitable to encrypt the plaintext information and it has good encryption effect.

TABLE III. INFORMATION ENTROPY

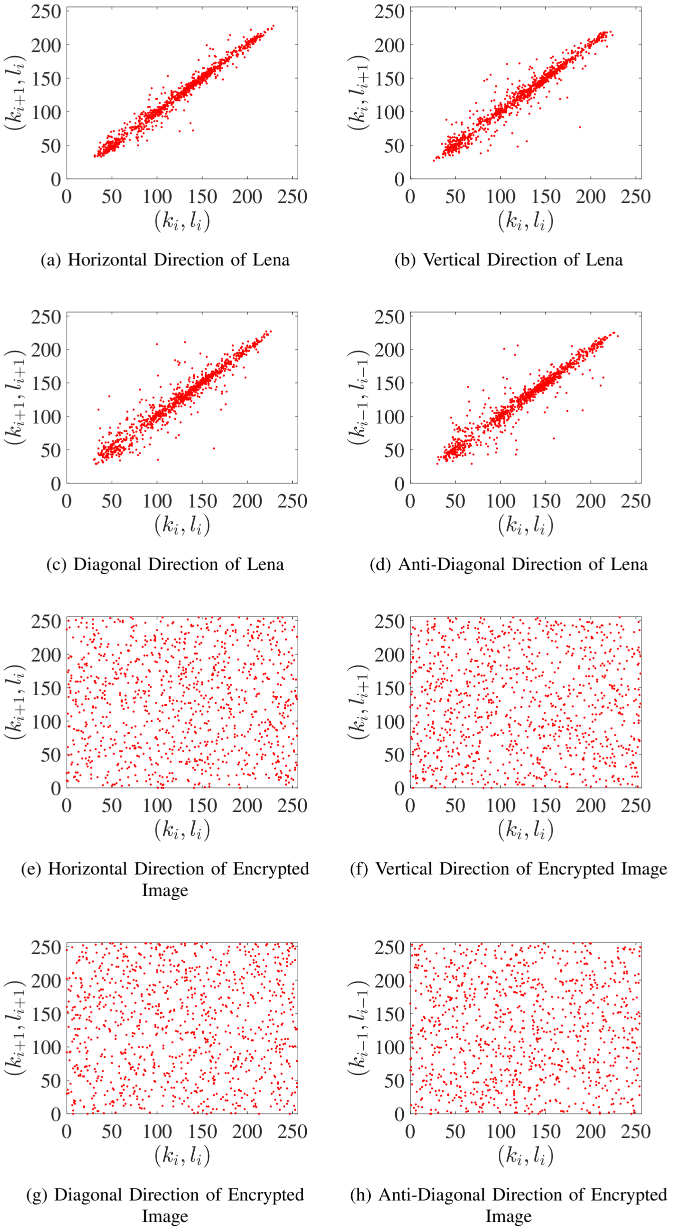
Image	Plaintext Image	Encrypted Image
Lena	7.4456	7.9915
Cameraman	7.0097	7.9902
Barbara	7.4664	7.9916
Lena in Ref. [32]	7.4456	7.9907
Lena in Ref. [33]	7.4456	7.9768

D. Analysis of Correlation Coefficient

The high correlation between the pixels of the plaintext image makes the image look clear and one may distinguish the image information. The correlation coefficient of unencrypted image is usually large, and the encryption algorithm will reduce the correlation between pixels to zero or close to zero. If N pairs of adjacent pixels are taken from the image and their gray value is (e_i, f_i) $(i = 1, 2, \cdots, N)$, the formula of correlation coefficient for vectors $\boldsymbol{e} = \{e_i\}$ and $\boldsymbol{f} = \{f_i\}$ is given as follows [34]:

$$\begin{cases} r_{\boldsymbol{ef}} = \dfrac{\operatorname{cov}(\boldsymbol{e}, \boldsymbol{f})}{\sqrt{D(\boldsymbol{e})}\sqrt{D(\boldsymbol{f})}} \\ \operatorname{cov}(\boldsymbol{e}, \boldsymbol{f}) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} (e_i - E(\boldsymbol{e}))(f_i - E(\boldsymbol{f})) \\ D(\boldsymbol{e}) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} (e_i - E(\boldsymbol{e}))^2, E(\boldsymbol{e}) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} e_i \end{cases}$$

If the e_i denotes the pixel value in position (k_i, l_i), and the f_i denotes the pixel value in position (k_{i+1}, l_i), then the calculation result is the correlation coefficient in the horizontal direction. Similarly, 1000 pairs of pixel points of Lena are randomly selected in the vertical, horizontal, diagonal and anti-diagonal directions, and the corresponding correlation coefficients are shown in Table IV. Meanwhile, the correlations of Lena and the encrypted images are given in Fig. 8, the correlation coefficient of the encrypted image of proposed algorithm has been close to 0.



(a) Horizontal Direction of Lena

(b) Vertical Direction of Lena

(c) Diagonal Direction of Lena

(d) Anti-Diagonal Direction of Lena

(e) Horizontal Direction of Encrypted Image

(f) Vertical Direction of Encrypted Image

(g) Diagonal Direction of Encrypted Image

(h) Anti-Diagonal Direction of Encrypted Image

Fig. 8. Correlation Analysis of Lena Image

E. Differential Analysis

A secure cryptographic system should be highly sensitive to small changes in the key or plaintext image during encryption

TABLE IV. CORRELATION COEFFICIENTS OF PLAINTEXT AND CIPHERTEXT

Image	Horizontal	Vertical	Diagonal	Average
Lena	0.9831	0.9737	0.9666	0.9745
Encrypted	−0.0092	−0.0116	0.0114	0.0086
Ref. [35]	0.0141	0.0296	0.0054	0.0164
Ref. [36]	−0.0253	0.0026	0.0091	0.0123

and decryption. Especially, if the small changes in the plaintext image will produce completely different encrypted images, then it will be more effectively to resist chosen-plaintext attacks. NPCR and UACI are often used to analyze whether the encryption algorithm has a good encryption effect and security. NPCR is the proportion of different pixel numbers in all pixel points. UACI represents the average difference of the pixel values of the encrypted image when the original image is with one pixel (or some pixels) different in pixel values. The calculation formulas are given as follows [37]:

$$\begin{cases} \text{NPCR} = \dfrac{\sum\limits_{i=1}^{M}\sum\limits_{j=1}^{N} D_{(i,j)}}{M \times N} \times 100\% \\ D_{(i,j)} = \begin{cases} 1, x\left(i,j\right) \neq x'\left(i,j\right) \\ 0, x\left(i,j\right) = x'\left(i,j\right) \end{cases} \\ \text{UACI} = \dfrac{\sum\limits_{i=1}^{M}\sum\limits_{j=1}^{N} \frac{x(i,j)-x'(i,j)}{255}}{M \times N} \times 100\% \end{cases}$$

where x is the encrypted image, and x' is the new encrypted image when the plaintext is changed by one pixel or some pixels. If the 97 in 99-th pixel value of Lena image is changed to 98, the 132 in 99-th pixel value of Barbara image is changed to 133, and the 156 in 99-th pixel value of Cameraman image is changed to 157, then the NPCR and UACI of corresponding encrypted images are shown in Table V, and they are very close to the expected values (NPCR ≈ 99.6094%, UACI ≈ 33.4635%). So the encryption algorithm can resist plaintext attacks and has significant encryption effect.

TABLE V. NPCR AND UACI OF ENCRYPTED IMAGE(%)

Image	NPCR	UACI
Lena	99.6174	33.5231
Cameraman	99.6323	33.3901
Barbara	99.6281	33.3537
Ref. [38]	99.8700	33.2900
Ref. [39]	99.2402	33.3873

F. NIST Test

NIST test is used to verify the random characteristics of random sequences, and it includes 15 tests, such as single bit frequency, longest-run-of-ones and non-overlapping template matching [40]. If the random sequence can pass all NIST test, i.e., the p-values are greater than 0.01, then the random sequence has good randomness. The pseudo-random sequences Z are obtained by the chaotic sequences X, and the results of NIST test for pseudo-random sequences Z are shown in Table VI, i.e., most tests of NIST are passed. Similarly, the results of NIST test for pseudo-random sequences W are shown in Table VII, and all the p-values are greater than 0.01, so the NIST test is passed.

VI. CONCLUSION

Through the combined controllers of trigonometric function, a class of asymptotically stable nominal linear systems are controlled to be hyperchaotic system, and a 11-dimensional hyperchaotic systems with 9 positive Lyapunov exponents is constructed. Meanwhile, an encryption algorithm of scrambling, sequential diffusion and inverse diffusion is designed based on the new hyperchaotic system. The encryption algorithm has many key parameters and initial values, and the key is related to plaintext information. To some extent, it has a large enough key space and can resist exhaustive attack and chosen-plaintext attack, etc. An example of image encryption is given by the simulation experiments, and it shows that the encryption algorithm based on higher-dimensional hyperchaotic system is feasible, effective and secure. Therefore, the chaos-based encryption algorithm can be applied to the secure communication in the near future, such as the encryption of images, video and other multimedia information.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. This research was supported by the National Natural Science Foundation of China (Grant No. 11847051), the Natural Science Foundation of Fujian Province of China (Grant Nos. 2022J01895 and 2019J05107), Fujian Key Laboratory of Granular Computing and Applications, Institute of Meteorological Big Data-Digital Fujian and Fujian Key Laboratory of Data Science and Statistics (Minnan Normal University), China.

REFERENCES

[1] K. M. Hosny, S. T. Kamal, and M. M. Darwish, "A color image encryption technique using block scrambling and chaos," *Multimedia Tools and Applications*, vol. 81, no. 1, pp. 505–525, 2022.

[2] Z. Madouri, N. H. Said, and A. A. Pacha, "Image encryption algorithm based on digital filters controlled by 2d robust chaotic map," *Optik*, vol. 264, p. 169382, 2022.

[3] K. Jain, A. Aji, and P. Krishnan, "Medical image encryption scheme using multiple chaotic maps," *Pattern Recognition Letters*, vol. 152, pp. 356–364, 2021.

[4] M. Hamdi, J. Miri, and B. Moalla, "Hybrid encryption algorithm (HEA) based on chaotic system," *Soft Computing*, vol. 25, no. 3, pp. 1847–1858, 2021.

[5] Y. Zhao and L. Liu, "A bit shift image encryption algorithm based on double chaotic systems," *Entropy*, vol. 23, no. 9, p. 1127, 2021.

[6] J. Chen, D. Yan, S. Duan, and L. Wang, "Memristor-based hyper-chaotic circuit for image encryption," *Chinese Physics B*, vol. 29, no. 11, p. 110504, 2020.

[7] D. S. Malik and T. Shah, "Color multiple image encryption scheme based on 3d-chaotic maps," *Mathematics and Computers in Simulation*, vol. 178, pp. 646–666, 2020.

[8] S. Liu, C. Li, and Q. Hu, "Cryptanalyzing two image encryption algorithms based on a first-order time-delay system," *IEEE MultiMedia*, vol. 29, no. 1, pp. 74–84, 2022.

TABLE VI. NIST TEST OF PSEUDO-RANDOM SEQUENCES Z

Test Items	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}
Frequency	0.9832	0.4348	0.6943	0.9944	0.2191	0.1228	0.1473	0.2878	0.1636	0.4208	0.5707
Frequency within a Block	0.8171	0.8927	0.7162	0.5393	0.4954	0.6526	0.9727	0.8223	0.3696	0.6062	0.2408
Runs	0.0677	0.8830	0.6384	0.3477	0.0061	0.9138	0.2216	0.2231	0.1433	0.3569	0.3886
Longest-Run-of-ones	0.8324	0.1244	0.3561	0.1394	0.9312	0.1289	0.6307	0.2157	0.6795	0.3849	0.4017
Binary matrix rank	0.0171	0.0000	0.0100	0.0171	0.0000	0.0000	0.0001	0.0000	0.0000	0.0555	0.0000
Discrete fourier transform	0.8618	0.7496	0.8846	0.9076	0.9076	0.6014	0.0519	0.1468	0.5617	0.5045	0.4333
Non-overlapping template matching	0.4893	0.7417	0.0216	0.3558	0.4551	0.8057	0.9046	0.2643	0.9285	0.6491	0.6057
Overlapping template matching	0.7772	0.4224	0.8261	0.2226	0.6515	0.2850	0.1263	0.3467	0.6426	0.1901	0.1616
Maurer's universal statistical	0.5107	0.7102	0.0458	0.4620	0.8034	0.1403	0.6694	0.8622	0.5631	0.1625	0.0324
Linear complexity	0.3866	0.7242	0.5818	0.4825	0.2231	0.2128	0.1809	0.7538	0.6493	0.7724	0.0120
Serial	0.6960	0.2418	0.3559	0.9198	0.0830	0.6367	0.1989	0.0726	0.0009	0.2406	0.0984
Approximate entropy	0.6596	0.8243	0.5361	0.1740	0.6640	0.9041	0.1823	0.5347	0.6192	0.2154	0.6175
Cumulative sums	0.8460	0.9963	0.5493	0.7691	0.7876	0.9185	0.9963	0.5493	1.0000	0.0807	0.2471
Random excursions	0.7036	0.3428	0.5153	0.5799	0.6196	0.6665	0.3548	0.1823	0.4403	0.8451	0.1190
Random excursions variant	0.4867	0.1584	0.2085	0.6523	0.5079	0.6144	0.6058	0.3360	0.1458	0.7315	0.0500

TABLE VII. NIST TEST OF PSEUDO-RANDOM SEQUENCES W

Test Items	W_1	W_2	W_3	W_4	W_5
Frequency	0.5035	0.2368	0.5452	0.8407	0.4139
Frequency within a Block	0.0971	0.0691	0.9898	0.9419	0.4977
Runs	0.2244	0.7435	0.9714	0.2807	0.4323
Longest-Run-of-ones	0.6785	0.2117	0.1600	0.5415	0.5751
Binary matrix rank	0.1087	0.2395	0.0560	0.0116	0.2666
Discrete fourier transform	0.5814	0.7277	0.4333	0.8390	0.5423
Non-overlapping template matching	0.0649	0.0876	0.8835	0.7137	0.7514
Overlapping template matching	0.4290	0.0828	0.2784	0.7073	0.9158
Maurer's universal statistical	0.9723	0.6845	0.2071	0.6439	0.1407
Linear complexity	0.9124	0.1711	0.8215	0.1285	0.2014
Serial	0.6145	0.7881	0.2658	0.1512	0.1737
Approximate entropy	0.9262	0.6077	0.7772	0.6742	0.3726
Cumulative sums	0.9998	0.5579	1.0000	0.7036	0.1673
Random excursions	0.6888	0.9516	0.9058	0.8177	0.7348
Random excursions variant	0.9562	0.6307	0.9703	0.9804	0.6885

[9] H. Fan, H. Lu, C. Zhang, M. Li, and Y. Liu, "Cryptanalysis of an image encryption algorithm based on random walk and hyperchaotic systems," *Entropy*, vol. 24, no. 1, p. 40, 2021.

[10] W. Liu, K. Sun, and S. He, "SF-SIMM high-dimensional hyperchaotic map and its performance analysis," *Nonlinear Dynamics*, vol. 89, no. 4, pp. 2521–2532, 2017.

[11] C. Shen, S. Yu, J. Lü, and G. Chen, "A systematic methodology for constructing hyperchaotic systems with multiple positive lyapunov exponents and circuit implementation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 3, pp. 854–864, 2013.

[12] C. Chen, K. Sun, and S. He, "A class of higher-dimensional hyperchaotic maps," *The European Physical Journal Plus*, vol. 134, no. 8, pp. 1–13, 2019.

[13] C. Shen, S. Yu, J. Lü, and G. Chen, "Constructing hyperchaotic systems at will," *International Journal of Circuit Theory and Applications*, vol. 43, no. 12, pp. 2039–2056, 2015.

[14] Z. Peng, W. Yu, J. Wang, Z. Zhou, J. Chen, and G. Zhong, "Secure communication based on microcontroller unit with a novel five-dimensional hyperchaotic system," *Arabian Journal for Science and Engineering*, vol. 47, no. 1, pp. 813–828, 2022.

[15] E. Yavuz, R. Yazıcı, M. C. Kasapbaşı, and E. Yamaç, "A chaos-based image encryption algorithm with simple logical functions," *Computers & Electrical Engineering*, vol. 54, pp. 471–483, 2016.

[16] F. Özkaynak and A. B. Özer, "Cryptanalysis of a new image encryption algorithm based on chaos," *Optik*, vol. 127, no. 13, pp. 5190–5192, 2016.

[17] X. Chai, Y. Chen, and L. Broyde, "A novel chaos-based image encryption algorithm using DNA sequence operations," *Optics and Lasers in engineering*, vol. 88, pp. 197–213, 2017.

[18] Ü. Çavuşoğlu, S. Kaçar, I. Pehlivan, and A. Zengin, "Secure image encryption algorithm design using a novel chaos based S-Box," *Chaos, Solitons & Fractals*, vol. 95, pp. 92–101, 2017.

[19] A. Y. Niyat, M. H. Moattar, and M. N. Torshiz, "Color image encryption

based on hybrid hyper-chaotic system and cellular automata," *Optics and Lasers in Engineering*, vol. 90, pp. 225–237, 2017.

[20] S. Zhu, C. Zhu, and W. Wang, "A new image encryption algorithm based on chaos and secure hash sha-256," *Entropy*, vol. 20, no. 9, p. 716, 2018.

[21] A. Gupta, D. Singh, and M. Kaur, "A novel image encryption using memetic differential expansion based modified logistic chaotic map," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 7, 2019.

[22] M. B. Farah, R. Guesmi, A. Kachouri, and M. Samet, "A novel chaos based optical image encryption using fractional fourier transform and dna sequence operation," *Optics & Laser Technology*, vol. 121, p. 105777, 2020.

[23] I. AlBidewi and N. Alromema, "Ultra-key space domain for image encryption using chaos-based approach with DNA sequence," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021.

[24] L. Chen, C. Li, and C. Li, "Security measurement of a medical communication scheme based on chaos and DNA coding," *Journal of Visual Communication and Image Representation*, vol. 83, p. 103424, 2022.

[25] M. I. Moussa, E. I. Abd El-Latif, and N. Majid, "Enhancing the security of digital image encryption using diagonalize multidimensional nonlinear chaotic system," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022.

[26] S. T. Liu and L. Zhang, "Surface chaos-based image encryption design," in *Surface Chaos and Its Applications*. Springer, 2022, pp. 321–346.

[27] S. Yu and G. Chen, "Anti-control of continuous-time dynamical systems," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 6, pp. 2617–2627, 2012.

[28] C. Shen, S. Yu, J. Lü, and G. Chen, "Designing hyperchaotic systems with any desired number of positive lyapunov exponents via a simple model," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 8, pp. 2380–2389, 2014.

[29] J. He and S. Yu, "Construction of higher-dimensional hyperchaotic systems with a maximum number of positive lyapunov exponents under average eigenvalue criteria," *Journal of Circuits, Systems and Computers*, vol. 28, no. 09, p. 1950151, 2019.

[30] S. El Assad and M. Farajallah, "A new chaos-based image encryption system," *Signal Processing: Image Communication*, vol. 41, pp. 144–157, 2016.

[31] C. E. Shannon, "Communication theory of secrecy systems," *The Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[32] V. Folifack Signing, T. Fozin Fonzin, M. Kountchou, J. Kengne, and Z. T. Njitacke, "Chaotic jerk system with hump structure for text and image encryption using dna coding," *Circuits, Systems, and Signal Processing*, vol. 40, no. 9, pp. 4370–4406, 2021.

[33] G. Kaur, R. Agarwal, and V. Patidar, "Color image encryption scheme based on fractional hartley transform and chaotic substitution–permutation," *The Visual Computer*, vol. 38, no. 3, pp. 1027–1050, 2022.

[34] G. Chen, Y. Mao, and C. K. Chui, "A symmetric image encryption scheme based on 3D chaotic cat maps," *Chaos, Solitons & Fractals*, vol. 21, no. 3, pp. 749–761, 2004.

[35] J. Gayathri and S. Subashini, "An efficient spatiotemporal chaotic image cipher with an improved scrambling algorithm driven by dynamic diffusion phase," *Information Sciences*, vol. 489, pp. 227–254, 2019.

[36] R. Guesmi and M. Farah, "A new efficient medical image cipher based on hybrid chaotic map and dna code," *Multimedia tools and applications*, vol. 80, no. 2, pp. 1925–1944, 2021.

[37] J. Cai and J. He, "A new hyperchaotic system generated by an external periodic excitation and its image encryption application," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 26, no. 3, pp. 418–430, 2022.

[38] P. T. Akkasaligar and S. Biradar, "Selective medical image encryption using dna cryptography," *Information Security Journal: A Global Perspective*, vol. 29, no. 2, pp. 91–101, 2020.

[39] P. N. Lone, D. Singh, and U. H. Mir, "A novel image encryption using random matrix affine cipher and the chaotic maps," *Journal of Modern Optics*, vol. 68, no. 10, pp. 507–521, 2021.

[40] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert *et al.*, "Nist special publication 800-22: A statistical test suite for the validation of random number generators and pseudo random number generators for cryptographic applications," *NIST Special Publication*, vol. 800, p. 22, 2010.

Exploring Regression-based Approach for Sound Event Detection in Noisy Environments

Soham Dinesh Tiwari
Department of Computer Science & Engineering
Manipal Institute of Technology
Manipal, India 576104

Karanth Shyam Subraya
Department of Computer Science & Engineering
Manipal Institute of Technology
Manipal, India 576104

***Abstract*—Sound-event detection enables machines to detect when a particular sound event has occurred in addition to classifying the type of event. Successful detection of various sound events is paramount in building secure surveillance systems and other smart home appliances. However, noisy events and environments exacerbate the performance of many sound event detection models, rendering them ineffective in real-world scenarios. Hence, the need for robust sound event detection algorithms in noisy environments with low inference times arises. You Only Hear Once (YOHO) is a purely convolutional architecture that uses a regression-based approach for sound-event-detection instead of the more common, frame-wise classification-based approach. The YOHO architecture proved robust in noisy environments, outperforming convolutional recurrent neural networks popular in sound event detection systems. Additionally, different ways to enhance the performance of the YOHO architecture are explored, experimenting with different computer vision architectures, dynamic convolutional layers, pretrained audio neural networks and data augmentation methods to help improve the performance of the models on noisy data. Amongst several modifications to the YOHO architecture, the Frequency Dynamic Convolution Layers helped improve the internal model data representations by enforcing frequency-dependent convolution operations, which helped improve YOHO performance on noisy audios in outdoor and vehicular environments. Similarly, the FilterAugment data augmentation method and Convolutional Block Attention Module helped improve YOHO's performance on the VOICe dataset containing noisy audios by augmenting the data and improving internal model representations of the input audio data using attention, respectively.**

Keywords*—Sound Event Detection (SED); sound event classification; frequency dynamic convolution; audio processing; FilterAugment; data augmentation; vision transformers; Pretrained Audio Neural Networks (PANN); Convolutional Block Attention Module (CBAM)

I. Introduction

In machine learning, sound event detection (SED) identifies the different sounds in an audio file and identifies the start and end time of a particular sound event in the audio. Various applications use SED, such as speech recognition, audio surveillance [1], and context-based indexing and data retrieval in a multimedia database [2].

Most of the research and development in SED today is focused on building sound event detection systems that can be trained using weakly labelled data, i.e., audios without timestamps for the occurrence of each event or unlabelled data [3]–[5]. Hence, there is an increased focus on using models that employ semi-supervised learning [4], [6], [7] to learn from weakly labelled and unlabelled data. Moreover, most of these works use sequential models or transformer architectures to leverage the sequential nature of audio data [8]–[10].

However, the consequence of using sequential and transformer-based architectures is increased model complexity, machine computation requirements, and inference times. In addition, the audios from various sources are seldom devoid of any interfering noise or disturbance in real-life. Consequently, this makes such models unsuitable for deployment in smart devices, which have constraints on the compute available and often are required to make inferences in real-time, in addition to often operating in noisy environments. Architectures with short inference times, low model parameters, and high accuracy enable smart devices to deliver accurate insights more quickly.

The You Only Hear Once (YOHO) architecture proposed by S. Venkatesh et al. draws inspiration from the famous computer vision architecture - You Only Look Once (YOLO) [11] and only makes use of different forms of convolutions, with no sequential layers. The YOHO algorithm matches the performance of the various state-of-the-art algorithms on datasets such as Music Speech Detection Dataset [12], TUT Sound Event [13], and Urban-SED datasets [14] and at lower inference times. The fast inference can be ascribed to YOHO's regression-based approach, which takes the entire audio stream as input and predicts each audio event's start and end time using regression instead of performing framewise classification.

Hence, this work tests the performance of the You Only Hear Once (YOHO) [15] algorithm on noisy audio data from different acoustic environments like indoors, outdoors, and in vehicles. The experiments show that the standard YOHO architecture is better than other popular sound event detection architectures when tackling noisy audios. Thus, we used the regression-based sound event detection approach from YOHO and combined it with other computer vision architectures, pretrained audio neural networks and methods like attention, data augmentation and dynamic convolutions to increase model performance in noisy environments. However, it was difficult to improve on the standard YOHO architecture's F1 scores. In the end, we were able to improve the performance of YOHO on the VOICe dataset by replacing the standard convolutional layers in the architecture with frequency dynamic convolution layers which helped mitigate translational invariance along the frequency axis of the log-Mel spectrograms. We also made use of the FilterAugment data augmentation method and Convolutional Block Attention Module (CBAM) to improve YOHO's

F1 scores for specific noise environments. Hence, in the end we modified the novel, purely convolutional, regression-based architecture proposed by S. Venkatesh et al. [15] to use attention, better internal data representations and data augmentation to further improve YOHO's SED performance on noisy data with a negligible increase in the model parameters.

The first section, "Introduction" explains the motivation for undertaking this work. The second chapter, "Background Theory and Literature Review", deals with the different concepts used in this work and discusses relevant research. The third chapter, "Methodology and Implementation Details", elaborates on the architecture designs and the steps undertaken in the experiments. The fourth chapter, "Results and Analysis" delineates and analyses the results. The fifth and ultimate chapter, "Conclusion and Future Work" provides a gist of the future scope of the research and possible technical improvements.

II. Background Theory and Literature Review

This section discusses the current state of research in sound event detection. It also elaborates on the different architectures and techniques in addition to the dataset and various attention and activation functions used in this work.

A. State of Research in SED

Early approaches for SED had adapted techniques from music information retrieval and speech recognition, such as Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs) [16]. However, HMMs could not deal with polyphonic audio containing co-occurring sound events.

The advent of deep learning demonstrated the adeptness of deep neural networks in multi-label sound event classification and dealing with polyphonic audio. While a simple feed-forward neural network (FFN) could perform multi-label classification, it was limited in its ability to model the temporal information in signals. Consequently, plain FNNs did not gain popularity for use in SED applications.

In 2017, Cakir et al. showed that the Convolutional Recurrent neural networks (CRNNs) [8] architectures were well suited for the SED task. The recurrent neural networks (RNNs) combined with convolutional neural networks (CNNs) could capture the audio's local and global context.

The following aspects helped the CRNN architecture perform well in the SED task:

- Using successive blocks of convolutional layers and non-linear transformations, the architecture learned distinguishing characteristics in the input log-Mel spectrograms and generated feature embeddings.
- The recurrent layers would then model the temporal dependencies in the feature embeddings from the convolutional layers.
- The framewise classification approach was used to predict the existence of audio events in each audio frame.

However, a drawback of the CRNN architecture was that processing times were higher, resulting from the sequential layers preventing the architecture from fully utilising GPUs' parallel computation.

To shorten inference durations without compromising high precision and recall, S. Venkatesh et al. introduced the You Only Hear Once (YOHO) architecture for SED [15]. The You Only Look Once (YOLO) algorithm, which is popular in Computer Vision, served as an inspiration for the YOHO algorithm. YOLO significantly reduced the processing durations for the input images and enabled real-time object recognition by changing the prediction of object bounding boxes from a classification problem to a regression problem. Similarly, YOHO transforms the frame-based, multi-label classification problem of acoustic boundary detection into a regression problem. To achieve this, multiple sets of three different output neurons were used, one for each class. The presence of an event class is detected by one neuron, and its start and endpoints are predicted by the other two in each set for the particular class. As a result, on numerous datasets, the YOHO article showed a higher F-measure and lower error rate than CRNNs [15].

The current focus of the majority of SED research and development is building sound event detection systems that can be trained using weakly labelled data (audios without timestamps for the occurrence of each event) or unlabelled data [3]–[5] As a result, semi-supervised learning models are being used more frequently [4], [6], [7] to learn from weakly labelled and unlabelled data. Additionally, the majority of these works take advantage of the sequential aspect of audio data by using sequential models or transformer architectures [7]–[9]. However, this work aims to explore the regression-based, supervised learning approach for SED; hence, semi-supervised learning approaches for SED are not a focus of this work.

B. VOICe Dataset

VOICe is a new dataset for developing and evaluating generalizable sound event detection and domain adaptation methods. VOICe is offered for sound event detection domain adaptation from one acoustic scene to another or between sound events with background and without background noise [17]. The VOICe dataset consists of 207 different audio mixtures containing audios with three different sound event labels:

1) babycry
2) gunshot
3) glassbreak

These audio events are then superimposed with audios from the following 3 acoustic scenes:

1) outdoor
2) indoor
3) vehicle

The noisy audios are mixed at 2 different sound-to-noise-ratio (SNR) levels:

1) -3 dB
2) -9 dB

The dataset also consists of 207 audio mixtures devoid of background noise. Consequently, there are a total of 1449 audio mixtures - 207 mixtures x 3 acoustic scenes x 2 SNRs

= 1242 noisy audios + 207 clean audios. These audio files are divided into “training”, “testing”, and “validation” sets. The testing and validation audio files are 60 seconds long, whereas the training audio files last about 180 seconds.

We used the noisier SNR -9 dB audios from the VOICe dataset [17] to assess the performance of YOHO and other architectures in this work on noisy audio. It is a dataset that has been artificially constructed utilising noise from various acoustic scenes and sound events from the TAU Urban Acoustic Scenes 2019 development set and TUT Rare Sound Events 2017 development set, respectively. We would have an algorithm that is resilient to noise and performs sound event recognition considerably faster than other contemporary algorithms if it can perform well on noisy data and provide scores at least matching the CRNN algorithm. Consequently, it would become a candidate model for portable sound event detection devices used in natural, noisy environments.

C. YOHO Algorithm

Instead of using the conventional, bigger YOLO architecture, S. Venkatesh et al. modified the final layers of the MobileNet [15] architecture to implement the YOHO architecture. As can be seen in Table I, YOHO is purely a convolutional neural network. The initial MobileNet architectural layers are retained in the initial half of the table. Layers adjusted for the target dataset are in the latter half.

Consequently, YOHO has faster inference times than designs with sequential layers like CRNNs because it is built as a solely convolutional neural network with no recurrent layers. The probability and the beginning and ending times of each event label are output from the neural network, which takes as input log-Mel spectrograms of the audios. In addition, because this end-to-end method directly predicts acoustic boundaries, it takes lesser time for post-processing and smoothing [18].

Log-Mel spectrograms are used as the YOHO model’s input. Next, a 3x3 2D convolution with a stride of 2 is performed after reshaping the input, halving the time and frequency dimensions. The MobileNet design uses depth-wise separable convolutions [19] with 3x3 filters, followed by point-wise convolutions [20] with 1x1 filters. All convolutions except the final layer were followed by batch normalisation [21] and ReLU activations [22]. Every time a stride of two is used, the time and frequency dimensions are halved.

D. PANNs: Large-scale Pretrained Audio Neural Networks for Audio Pattern Recognition

It is common practice in computer vision and natural language processing for systems to be pretrained on large-scale datasets. The pretrained systems then generalise well to several downstream tasks. However, the research on building pretrained neural networks trained on large-scale audio datasets is limited. Pretrained audio neural networks (PANNs) [23] are trained on the large-scale AudioSet dataset [24]. These PANNs are transferred to other downstream audio-related tasks. The best PANN system achieved mean average precision (mAP) of 0.439 on AudioSet tagging, outperforming the previous state-of-the-art result of 0.392 [23].

E. ViT: Vision Transformer

The Vision Transformer, or ViT [25], is a transformer-based model architecture for image classification. Each image is split into fixed-size patches. The patches are then converted to linear embeddings using a linear transformation. The linear embeddings are then added to the positional embedding, and the resulting sequence of vectors is fed to a standard Transformer encoder. Finally, an extra learnable “classification token” in the output sequence is used to perform image classification [25].

F. CoAtNet: Marrying convolution and Attention for All Data Sizes

Vanilla Vision Transformers lack the inductive biases that traditional convolutional networks possess. However, Vision transformers have the advantage of utilising attention over their input [26]. Hence, a model which uses convolution and attention in machine learning benefits from two fundamental aspects – higher generalisation and higher model capacity. Convolutional layers have better generalisation, while attention in the transformer layer yields higher model capacity. Hence CoAtNet [26] is a hybrid model based on two key insights:

1) By using simple relative attention, depth-wise convolution and self-attention can be naturally fused.
2) By stacking convolution layers and attention layers in a principled manner, generalisation, capacity, and efficiency are dramatically improved.

CoAtNet has the best of both convolution networks and Transformers. CoAtNet not only has the generalisation ability of convolution networks because of favourable inductive biases but also has the advantage of superior scalability from transformers resulting in faster convergence and improved efficiency [26].

G. KNNs: Kervolutional Neural Networks

KNNs [27] are an effort to establish convolution in non-linear space. Existing CNN architectures primarily leverage activation layers which only provide point-wise non-linearity. Kervolution (kernel convolution) was introduced to approximate complex behaviours of human perception systems by leveraging the kernel trick. It claims that using the kernel trick helps it generalise convolution, capture higher-order interactions of features via patch-wise kernel functions, and enhance the model capacity without introducing additional parameters. Extensive experiments showed that kervolutional neural networks (KNN) achieve higher accuracy and faster convergence than baseline CNN [27].

H. CBAM: Convolutional Block Attention Module

Convolutional Block Attention Module (CBAM) [28] is an attention module designed for convolutional neural networks. The module sequentially computes attention maps along the channel and spatial dimensions given an intermediate feature map. The input feature maps are then multiplied by the attention maps to refine the adaptive features [28].

I. Frequency Dynamic Convolutions

When a pattern moves along the time axis in the time-frequency domain, it sounds the same because the frequency components are the same, but the timing is different. On the other hand, because the frequency component that makes up the acoustic properties of the sound event changes as it moves along the frequency axis, it would sound different [29]. The Frequency Dynamic Convolutions paper [29] highlights that due to its nature, the standard 2D convolution operation wrongly enforces translational invariance along both the time and frequency dimensions of an input log-Mel spectrogram. In contrast, log-Mel spectrograms exhibit invariance only along the time axis but not the frequency axis.

Frequency dynamic convolution [29] uses a frequency-adaptive kernel to enforce frequency-dependency on 2D convolution and improve the physical consistency of the model with sound events' time-frequency patterns. The inputs have three axes - the frequency axes, the time axes and the parallel channels. From the input, it first derives frequency-adaptive attention weights. Average pooling along the time axis is followed by two 1D convolution layers which are applied along the channel axis the input. 1D convolution was used rather than fully-connected (FC) layers to take into account neighbouring frequency components. Batch normalisation and ReLU are applied between two 1D convolution layers. The channel dimension is compressed into the number of basis kernels using 1D convolution layers. The softmax activation function is then used to normalise the values of the frequency-adaptive attention weights between zero and one.

Additionally, Softmax sets each frequency bin's weight sum to one. A temperature of 31 was applied to the softmax to achieve stable training and homogeneous learning of basis kernels. Then, using frequency-adaptive attention weights, a weighted sum of basis kernels is used to produce frequency-adaptive convolution kernels. The obtained kernel is used for frequency dynamic convolution operation just like a regular 2D convolution.

J. FilterAugment Data Augmentation

FilterAugment [30] is an improved version of frequency masking compared to SpecAugment [31]. SpecAugment involved simply masking a small time and frequency range of Mel spectrograms. While the simplicity of the time and frequency masking makes it easily adoptable during model training, they are unforgiving as they completely remove portions of information from the spectrogram.

FilterAugment was proposed to regularize acoustic models over various acoustic environments by mimicking acoustic filters. It approximates acoustic filters by applying random weights on randomly determined frequency bands, i.e., randomly increasing or decreasing the energy of these random frequency ranges of the log Mel spectrograms.

Hence, FilterAugment is an effective regularization approach for acoustic models as it extracts sound information from a wider range of frequencies and it proved that generalising the SED model over a broader frequency range enhances SED performance by a significant margin.

K. Mish Activation Function

$$f(x) = x \cdot tanh(softplus(x)) \tag{1}$$

$$softplus(x) = ln(1 + e^x) \tag{2}$$

The Mish activation function outperforms ReLU [22]. The advantages of the Mish activation function are: It has unbounded upper limit and bounded lower limit. It is a non-monotonic, self-regularized function and a self-gated function. It is continuously differentiable with infinite order. It falls under the class of $C\infty$. In contrast, ReLU falls under the class of $C0$. However, a disadvantage of Mish is that it is computationally expensive [32].

L. SERF Activation Function

Serf, or Log-Softplus ERror activation Function, is a type of activation function which is self-regularized and nonmonotonic in nature.

$$f(x) = x \cdot erf(softplus(x)) \tag{3}$$

erf is the error function. SERF outperforms Mish and ReLU on a variety of tasks like image classification, object detection, machine translation and sentiment classification on multiple datasets [33].

III. Methodology and Implementation Details

The following section describes the different modification made to each architecture for the downstream tasks as well as the different hyperparameters set for different aspects of model training. The codes have been made available on Github[1].

A. Audio Processing

The following steps are performed to process the audio files and corresponding annotations:

1) Convert stereo audio with SNR -9 dB into mono audio.
2) Read mono audio at a sample rate of 44,100 Hz, then segment it into windows, where the window length is set to 2.56s and hop length is set to 1.96s.
3) Generate model compatible output format for each window's annotations.
4) Generate log-Mel spectrograms for each window using the following parameters:
    a) Number of Mel bins = 40
    b) Number of Fast Fourier Transform components = 2048
    c) Window length for generating spectrograms = 1764 sample points
    d) Hop length for generating spectrograms = 441 sample points
    e) Minimum frequency $fmin$ of 0 Hz
    f) Maximum frequency $fmax$ of 22,050 Hz
5) Save the log-Mel spectrograms and the model compatible outputs using the $.npy$ file format for use during model training and validation.

[1] https://github.com/sohamtiwari3120/YOHO-on-VOICe

6) At the time of model training, FilterAugment [30] is optionally used to augment the existing dataset. It randomly increases or decreases the energy of random frequency regions of the log-Mel spectrogram.

B. Model Training Settings

The random seed was fixed at 0. The batch size for training the model was set at 32. We used the -9 dB audios from the VOICe dataset for training.

While training, the Adam [34] optimiser with the default learning rate set to 10^{-3}, epsilon value set to 10^{-7} and default weight decay set to 0. Moreover, a call-back was used to reduce the learning rate of the model by half during training whenever the validation loss during training did not decrease for 5 epochs. Gradient clipping was used to clip/keep the global norm values of the gradients less than or equal to 0.5.

The number of epochs was not fixed, and Early Stopping [35] regularization was utilised to stop model training if the validation loss would not decrease for 10 or more epochs.

C. Working Changes to YOHO Architecture

Each audio file is first segmented into overlapping windows with a window length of 2.56s and a hop length of 1.96s. The model input dimensions are determined by the length of the audio example and the number of Mel bins. For the VOICe dataset, the input log-Mel spectrogram shape is (257, 40), where 257 is the number of time steps and 40 is the number of Mel bins.

As shown in Table I, the output of the last 2D convolution layer is reshaped by flattening the last two dimensions. The final 1D convolution layer consisting of nine filters of unit length yields an output matrix of shape 9x9. The first dimension of the output corresponds to the time axis, and the second dimension corresponds to three neurons for each of the three classes in the VOICe dataset. The input spectrogram is divided into nine bins along the time axis. Every row in the output matrix represents the audio events' start and end times in the input spectrogram's corresponding bin.

As seen in Fig. 1, the first, fourth and seventh neurons perform binary classification at each time step to detect the presence of the respective audio events. The second and third neurons use regression to predict the start and end times of the first audio-event class. Similarly, for the fifth, sixth, eighth and ninth neurons and their respective second and third audio-event classes.

Table I describes the YOHO architecture for the VOICe dataset. Each convolution layer in the architecture makes use of 'same' padding to keep the shape of the output of the convolution layer same as that of its input.

D. Working Changes to CoAtNet + CBAM Architecture

The CoAtNet architecture expects input images which have equal height and width and three channels. Since the size of the log-Mel spectrograms of the VOICe dataset was (257, 40) , i.e., (number of time frames, number of mel bins), a learnable transposed convolution layer [36] is included before the CoAtNet architecture. The transposed convolution layer would then reshape the input spectrogram to (257, 257) followed by a 2D convolutional layer which would increase the number of channels to 3. To this transformed input CBAM attention (with reduction factor of 2 and kernel size of 3) is applied before passing it onto the CoAtNet architecture. The "CCTT" variant [26] of CoAtNet architecture is used, i.e., two MobileNet Convolution Layers along with two Transformer layers. Finally, the output of the CoAtNet architecture is passed through a 1D Convolution layer to increase the number of channels to 9. The changes are described in Table II.

TABLE I. YOHO'S ARCHITECTURE MODIFIED FOR VOICE DATASET

Layer type	Filters	Kernel Shape; Stride	Output shape
Reshape	-	-	(257, 40, 1)
Conv2D	32	(3, 3); 2	(129, 20, 32)
Conv2D-dw	-	(3, 3); 1	(129, 20, 32)
Conv2D	64	(1, 1); 1	(129, 20, 64)
Conv2D-dw	-	(3, 3); 2	(65, 10, 64)
Conv2D	128	(1, 1); 1	(65, 10, 128)
Conv2D-dw	-	(3, 3); 1	(65, 10, 128)
Conv2D	128	(1, 1); 1	(65, 10, 128)
Conv2D-dw	-	(3, 3); 2	(33, 5, 128)
Conv2D 256	256	(1, 1); 1	(33, 5, 256)
Conv2D-dw	-	(3, 3); 1	(33, 5, 256)
Conv2D 256	256	(1, 1); 1	(33, 5, 256)
Conv2D-dw	-	(3, 3); 2	(17, 3, 256)
Conv2D	512	(1, 1); 1	(17, 3, 512)

5x (Conv2D-dw;
Conv2D)
	-
	512
		(3, 3);
		(1, 1);
			(17, 3, 512)
			(17, 3, 512)

Conv2D-dw	-	(3, 3); 2	(9, 2, 512)
Conv2D	1024	(1, 1); 1	(9, 2, 1024)
Conv2D-dw	-	(3, 3); 1	(9, 2, 1024)
Conv2D	1024	(1, 1); 1	(9, 2, 1024)
Conv2D-dw	-	(3, 3); 1	(9, 2, 1024)
Conv2D	512	(1, 1); 1	(9, 2, 512)
Conv2D-dw	-	(3, 3); 2	(9, 2, 512)
Conv2D	256	(1, 1); 1	(9, 2, 256)
Conv2D-dw	-	(3, 3); 1	(9, 2, 256)
Conv2D	128	(1, 1); 1	(9, 2, 128)
Reshape	-	-	(9, 256)
Conv1D	9	(1); 1	(9, 9)

TABLE II. WORKING CHANGES TO COATNET + CBAM ARCHITECTURE

Name	Type	Output Shape
input	log-Mel spectrogram	(1, 257, 40)
make_input_square	ConvTranspose2d(...)	(1, 257, 257)
increase_channels_to_3	Conv2d(...)	(3, 256, 256)
cbam	CBAMBlock(...)	(3, 256, 256)
cn	CoAtNet(...)	(1, 9)
increase_1d_channels	Conv1d(...)	(9, 9)

E. Working Changes to ViT Architecture

The Vision Transformer (ViT) expects input images with three channels and height and width of even values since it splits each image into fixed-size patches. In our code, the patch size for ViT was set to 8. The depth of the transformer layer was set to 6, the number of heads to 16, the dimension of transformer tensors to 1024 and that of the feed-forward neural network layer to 2048.

Table III shows that the input log-Mel spectrogram is passed through a 2D Convolutional layer. This layer increases the number of channels to 3 and reduces the height of the image to an even value. The ViT layer outputs a 1D vector of 1024 values and 161 channels. Consequently, the output is passed through two successive 1D convolutional layers to obtain the final output in the desired shape (9, 9).

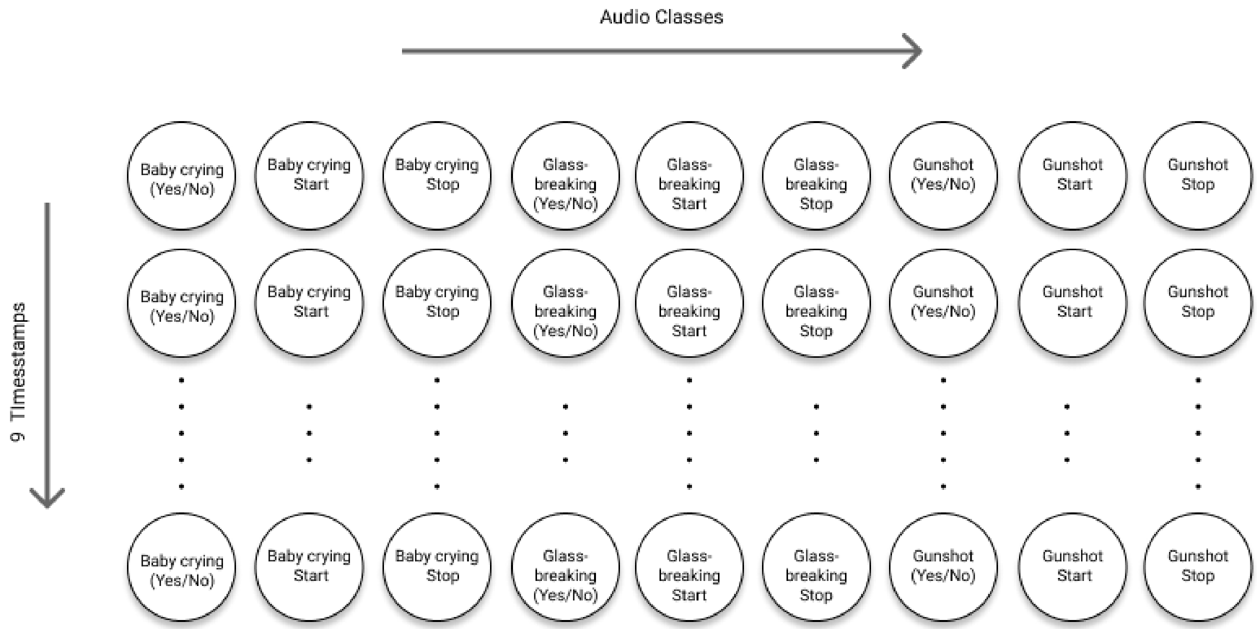


Fig. 1. Visualisation of Output Layer of YOHO for VOICe Dataset.

TABLE III. WORKING CHANGES TO VIT ARCHITECTURE

Name	Type	Output Shape
input	log-Mel spectrogram	(1, 257, 40)
increase_channels_to_3_and_reduce_height	Conv2d(...)	(3, 256, 40)
ViT	ViT(...)	(161, 1024)
head.0	Conv1d(...)	(9, 512)
head.1	Conv1d(...)	(9, 9)

F. Working Changes to YOHO + KNN Architecture

The YOHO architecture was modified by replacing every standard 2D convolution layer in the architecture by a 2D kervolution layer with a linear kernel.

G. Working Changes to YOHO + PANN Architecture

The CNN10 [23] variant of the PANN architecture expects log-Mel spectrograms with 64 Mel bins as input. Hence as described in Table IV, the input is processed and made compatible with PANN. PANN's output is then passed through a couple of transpose convolution layers followed by a couple of convolution layers to make the output compatible for YOHO. During training and inference, the weights of PANN pretrained on the large-scale AudioSet dataset were used for the CNN10 architecture.

TABLE IV. WORKING CHANGES TO YOHO + PANN ARCHITECTURE

Name	Type	Output Shape
input	log-Mel spectrogram	(1, 257, 40)
transpose	input.transpose(0, 2)	(40, 257, 1)
increase_channels_to_64	Conv2d(...)	(64, 257, 1)
transpose	input.transpose(0, 2)	(1, 257, 64)
PANN	CNN10(...)	(512, 16, 4)
transpose_conv2d_1	ConvTranspose2d(...)	(256, 16, 40)
transpose_conv2d_2	ConvTranspose2d(...)	(128, 257, 40)
reduce_channels_1	Conv2d(...)	(64, 257, 40)
reduce_channels_2	Conv2d(...)	(1, 257, 40)
YOHO	YOHO(...)	(9, 9)

H. Working Changes to YOHO + CBAM Architecture

The YOHO architecture is modified by inserting a Convolutional Block Attention Module (CBAM) before the input to every 2D depth wise convolution layer. The reduction factor was set to 2 and the kernel size set to 3 for every CBAM layer. The number of channels for each cbam layer was set to the number of channels of the output from the previous layer.

I. Working Changes to YOHO + Custom Rectangular Kernel Architecture

In an attempt to mitigate the problem of convolutional neural networks wrongly enforcing translational invariance along the frequency axis of the log-Mel spectrograms, a custom convolutional layer with "rectangular kernels" was applied before being input to the YOHO architecture. Fig. 2 shows that the layer consisted of 4 parallel 2D convolution layers:

1) Conv2D layer with filter size (3, 3)
2) Conv2D layer with filter size (log-Mel height//4, 3)
3) Conv2D layer with filter size (log-Mel height//2, 3)
4) Conv2D layer with filter size (log-Mel height*3//4,3)

J. Working Changes to YOHO + FilterAugment Architecture

The YOHO architecture remains the same. The difference is during training, FilterAugment data augmentation is used. The decibel range was set to (-6, 6), the band number range to (3, 6), minimum bandwidth to 6 and a linear filter type was used. The same set of hyperparameters are used for all experiments using FilterAugment data augmentation.

K. Working Changes to YOHO + FDY Architecture

The YOHO architecture was modified by replacing every standard 2D convolutional layer in the architecture by a 2D frequency dynamic convolution layer with the number of basis

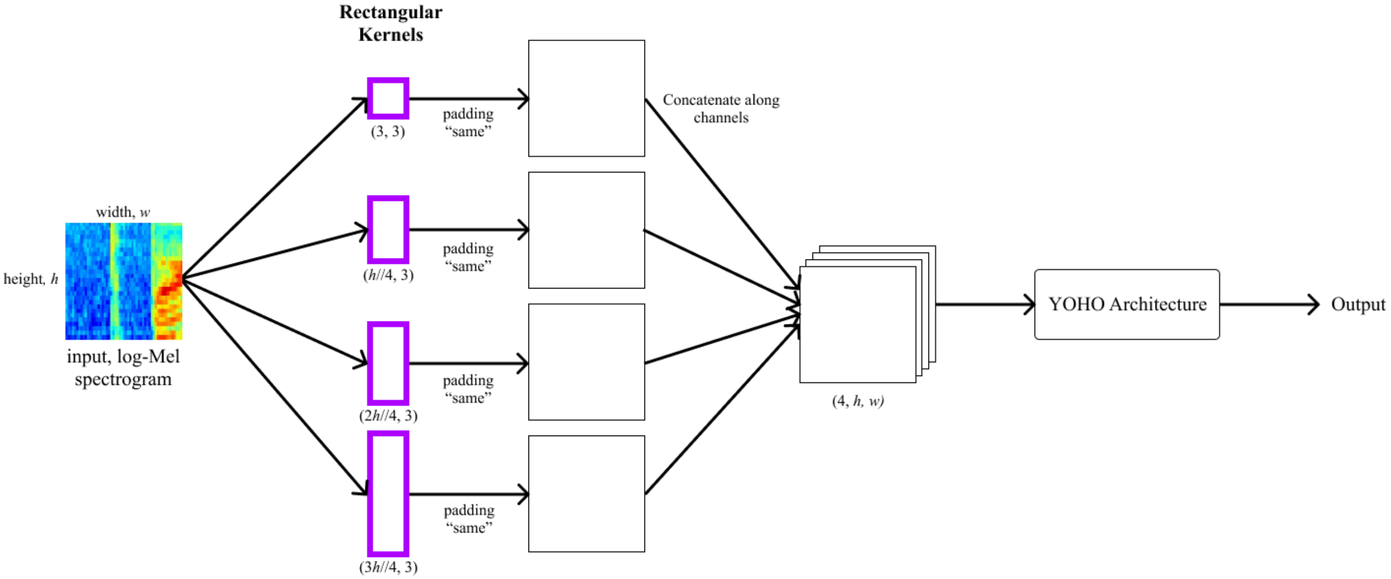


Fig. 2. Visualisation of Rectangular Kernel Layer Preceeding YOHO.

kernels set to 4 and the temperature set to 31 in every such layer. The same hyperparameters are used in all experiments using frequency dynamic convolutions.

L. Working Changes to YOHO + FDY + FilterAugment Architecture

The architecture is same as that described in the previous section, the difference being that the FilterAugment data augmentation method was being used during the training of the above architecture.

M. Working Changes to YOHO + FDY + FilterAugment + CBAM Architecture

The architecture is same as that described in the previous section, the difference being that the CBAM attention module (with reduction factor of 2 and kernel size of 3) is inserted before the input to every depth wise frequency dynamic convolutional layer.

IV. Results and Analysis

The results of all the experiments have been tabulated in Table V. More detailed results have been made available on Weights and Biases[2].

The results illustrate the following points:

1) YOHO, with an average F1 score of 0.8738, outperforms the CRNN architecture, which has an average F1 score of 0.7603, as given in [17].
2) The larger and more complex architectures, such as CoAtNet, ViT and PANN (with YOHO) with average F1 scores of 0.7589, 0.7966 and 0.8594 respectively are unable to match standard YOHO architecture's 0.8738 F1 score. This is surprising as these architectures are currently one of the best performing models in the multiple applications of Computer Vision and audio processing.
3) On average, the best performing architecture was the YOHO + CBAM architecture, with an average F1 score of 0.874. This architecture entailed adding attention to YOHO, which benefits from the inductive biases of a convolutional network. This is again surprising as the CoAtNet architecture, which combines the global attention of transformers and the inductive biases of the convolutional networks could not perform as well.
4) The proposed rectangular kernels layer (average F1 score of 0.8676), a simplistic approach to improve internal model representation of audio log-Mel spectrograms performed better than larger architectures like CoAtNet (average F1 score 0.7589), CRNN (average F1 score 0.7603), ViT (average F1 score 0.7966) and YOHO + PANN (average F1 score 0.8594) on the dataset.
5) YOHO modified with Kervolutional layers reported the lowest F1 scores in the Outdoor - 0.7883 and Vehicle - 0.8167 noise environments compared to all other variants of YOHO.
6) YOHO + FilterAugment reported the best F1 scores in the outdoor and the vehicle noise environments - 0.879 and 0.8918, respectively. In contrast, it performed the worst out of all architectures in the indoor noise environment, with an F1 score of 0.3871.
7) YOHO + FDY + FilterAugment reported very high scores in the outdoor and noise environments - 0.8746 and 0.8903, respectively. Moreover, frequency dynamic convolution layers seemed to counteract to some extent the drop in performance when using FilterAugment for indoor audios with an F1 score of 0.7681.
8) Combining YOHO + FDY + FilterAugment + CBAM did not perform as expected. Its F1 scores (0.8668, 0.8815 and 0.4534) are lesser than its individual

[2] https://wandb.ai/sohamtiwari3120/YOHO-on-VOICe?workspace=user-sohamtiwari3120

TABLE V. RESULTS OF ALL EXPERIMENTS

Architecture	**Outdoor F1**	**Vehicle F1**	**Indoor F1**	**Average F1**
CoAtNet	0.7566	0.7719	0.7483	0.7589
CRNN	0.7500	0.8004	0.7305	0.7603
ViT	0.7902	0.8223	0.7774	0.7966
YOHO	0.8714	0.8884	**0.8615**	0.8738
YOHO + KNN	0.7883	0.8167	0.7732	0.7927
YOHO + PANN (cnn10)	0.8598	0.8712	0.8472	0.8594
YOHO + Rectangular Kernels	0.8670	0.8792	0.8567	0.8676
YOHO + CBAM	0.8724	0.8905	0.8591	**0.8740**
YOHO + FDY	0.8674	0.8864	0.8607	0.8715
YOHO + Filt_Aug	**0.8790**	**0.8918**	*0.3871*	0.7193
YOHO + FDY + FILT_AUG	0.8746	0.8903	0.7681	0.8443
YOHO + FDY + FILT_AUG + CBAM	0.8668	0.8815	*0.4534*	0.7339

components - YOHO, YOHO + CBAM, YOHO + FilterAugment and YOHO + FDY + FilterAugment.

Table V shows that the standard YOHO architecture is adept at handling noisy audios, and few architectures and other variants of YOHO could improve upon the standard architecture's scores. It seems that complex architectures with a large number of parameters performed worse than standard YOHO on the VOICe dataset. One reason could be less training data, which large architectures usually require. The VOICe dataset could be sufficient for YOHO but insufficient for the more complex architectures.

Furthermore, architectures with sequential layers and transformers like CRNN and ViT performed worse than the purely convolutional YOHO architecture and its variants. Purely convolutional architectures can better utilise GPUs and hence train faster and longer. In addition, convolutional neural networks possess inductive biases, which help them generalise to unseen datasets. Hence, another reason sequential layer-based architectures did not perform well on noisy audios could be the lack of inductive biases and poor generalizability compared to YOHO. However, the reason why CoAtNet, which combines the inductive biases of convolutional layers and the attention of transformers, performed poorly on the dataset is unknown. The poor performance of CoAtNet is confounding, especially since YOHO combined with attention in the form of CBAM performed very well on audios from all three noise environments.

On the other hand, the experiments using convolutional block attention module (CBAM), FilterAugment and Frequency Dynamic Convolution (FDY) layers hint that improving internal model representations of the input audio data and data augmentation hold the key to achieving robust sound event detection on noisy audios.

V. CONCLUSION AND FUTURE WORK

The standard YOHO architecture is resilient to noise in audios, and its lightweight architecture with low inference times makes it a candidate model for use in portable sound event detection applications. Furthermore, this work found that combining the YOHO architecture with Frequency Dynamic Convolutions, Convolutional Block Attention Module, and FilterAugment data augmentation, helped obtain high F1 scores in specific noisy environments. This increase in performance can be attributed to improved internal model representations of the input audios with the help of data augmentation, which entailed a negligible increase in the number of parameters and model inference times.

The results suggest that improving audio representations in the model and data augmentation could help obtain better performance of SED models in noisy environments. Hence, future research can explore better representations for audios and research ways to mitigate the translational invariance along the frequency axis enforced by standard convolutional neural networks. This could involve improving the proposed rectangular kernels layer. Additionally, the models could be trained with representations from Google LEAF [37], which was one of the experiments we wanted to conduct. However, at the time of writing this paper, we could not find links to pretrained weights for the Google LEAF architecture and our compute and data were inadequate to train the architecture from scratch.

Another possible avenue of research could be to determine why the FilterAugment data augmentation method adversely affects only the audios containing noise from indoor environments while showing the opposite, helpful effects in audios containing vehicular and outdoor noise. This study could help unlock insights about noise in indoor environments and could help in the development of intelligent home audio devices.

Another direction of research would be to first train the larger and more complex architectures with more data to try and determine the ceiling for their performance in noisy environments. Furthermore, these experiments could help assess whether their performance can be improved using additional data or other factors that govern the model performance.

Finally, to determine why YOHO with convolutional block attention module performs so well and why CoAtNet does not. This could help us understand why CoAtNet's architecture, which benefits from inductive biases of convolutional networks and the global attention mechanism of transformer architectures, did not perform well in audios from noisy environments.

This work showed that the promising new regression-based, purely convolutional sound event detection architecture called You Only Hear Once can be effectively used in noisy environments. Furthermore, the regression-based approach was experimented with and used with other architectures and methods to find ways to better improve robustness to noise. The results indicate that better audio representations for the model and data augmentation techniques can help boost the performance of SED systems. Hopefully, this work will encourage more research in developing better audio representations and making audio-related models more robust to noise.

Acknowledgment

We want to thank the faculty of the Computer Science Department at Manipal Institute of Technology, Manipal, India, for their continued support and guidance. Special appreciation for Dr. Manjunath Mulimani (Email manjunath.mulimani@tuni.fi), who, during his tenure at Manipal Institute of Technology, guided us and helped us navigate various technical difficulties and advised possible research directions throughout the lifetime of this work. The authors would also like to thank Dr. Chng Eng Siong, Andrew Koh and Tanmay Khandelwal, part of the Speech Lab at Nanyang Technological University, Singapore.

References

[1] S. Ntalampiras, "Audio surveillance." [Online]. Available: www.witpress.com,

[2] "Sound Event Detection - Toni Heittola." [Online]. Available: https://homepages.tuni.fi/toni.heittola/research-sound-event-detection

[3] "Sound Event Detection in Domestic Environments - DCASE." [Online]. Available: https://dcase.community/challenge2022/task-sound-event-detection-in-domestic-environments

[4] H. Dinkel, X. Cai, Z. Yan, Y. Wang, J. Zhang, and Y. Wang, "Detection and Classification of Acoustic Scenes and Events 2021 A LIGHTWEIGHT APPROACH FOR SEMI-SUPERVISED SOUND EVENT DETECTION WITH UNSUPERVISED DATA AUGMENTATION." [Online]. Available: https://github.com/pytorch/pytorch

[5] A. Cheung, Q. Tang, C.-C. Kao, M. Sun, and C. Wang, "Detection and Classification of Acoustic Scenes and Events 2021 IMPROVED STUDENT MODEL TRAINING FOR ACOUSTIC EVENT DETECTION MODELS."

[6] R. Serizel, N. Turpault, H. Eghbal-Zadeh, and A. P. Shah, "Large-Scale Weakly Labeled Semi-Supervised Sound Event Detection in Domestic Environments," 7 2018. [Online]. Available: https://arxiv.org/abs/1807.10501v1

[7] K. Miyazaki, T. Komatsu, T. Hayashi, S. Watanabe, T. Toda, and K. Takeda, "Detection and Classification of Acoustic Scenes and Events 2020 CONFORMER-BASED SOUND EVENT DETECTION WITH SEMI-SUPERVISED LEARNING AND DATA AUGMENTATION."

[8] E. Cakir, G. Parascandolo, T. Heittola, H. Huttunen, and T. Virtanen, "Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, no. 6, pp. 1291–1303, 6 2017.

[9] D. De Benito-Gorron, D. Ramos, and D. T. Toledano, "A Multi-Resolution CRNN-Based Approach for Semi-Supervised Sound Event Detection in DCASE 2020 Challenge," *IEEE Access*, vol. 9, pp. 89 029–89 042, 2021.

[10] K. Miyazaki, T. Komatsu, T. Hayashi, S. Watanabe, T. Toda, and K. Takeda, "Detection and Classification of Acoustic Scenes and Events 2020 CONVOLUTION-AUGMENTED TRANSFORMER FOR SEMI-SUPERVISED SOUND EVENT DETECTION Technical Report."

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 779–788, 6 2015. [Online]. Available: https://arxiv.org/abs/1506.02640v5

[12] "2018:Music and/or Speech Detection - MIREX Wiki." [Online]. Available: https://music-ir.org/mirex/wiki/2018:Music_and/or_Speech_Detection

[13] "Acoustic scene classification - DCASE." [Online]. Available: http://dcase.community/challenge2017/task-acoustic-scene-classification

[14] "URBAN-SED - Home." [Online]. Available: http://urbansed.weebly.com/

[15] S. Venkatesh, D. Moffat, and E. Reck Miranda, "You Only Hear Once: A YOLO-like Algorithm for Audio Segmentation and Sound Event Detection." [Online]. Available: https://github.com/satvik-venkatesh/you-only-hear-once

[16] A. Mesaros, T. Heittola, T. Virtanen, and M. D. Plumbley, "Sound Event Detection: A Tutorial."

[17] S. Gharib, K. Drossos, E. Fagerlund, and T. Virtanen, "VOICe: A Sound Event Detection Dataset For Generalizable Domain Adaptation." [Online]. Available: https://doi.org/10.5281/zenodo.3514950

[18] S. Tiwari, K. Lakhotia, and M. Mulimani, "Evaluating robustness of You Only Hear Once(YOHO) Algorithm on noisy audios in the VOICe Dataset," 11 2021. [Online]. Available: https://arxiv.org/abs/2111.01205v1

[19] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 1800–1807, 10 2016. [Online]. Available: https://arxiv.org/abs/1610.02357v3

[20] "Pointwise Convolution Explained — Papers With Code." [Online]. Available: https://paperswithcode.com/method/pointwise-convolution

[21] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *32nd International Conference on Machine Learning, ICML 2015*, vol. 1, pp. 448–456, 2 2015. [Online]. Available: https://arxiv.org/abs/1502.03167v3

[22] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," 3 2018. [Online]. Available: https://arxiv.org/abs/1803.08375v2

[23] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 28, pp. 2880–2894, 12 2019. [Online]. Available: https://arxiv.org/abs/1912.10211v5

[24] J. F. Gemmeke, D. P. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio Set: An ontology and human-labeled dataset for audio events," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 776–780, 6 2017.

[25] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," 10 2020. [Online]. Available: https://arxiv.org/abs/2010.11929v2

[26] Z. Dai, H. Liu, Q. V. Le, and M. Tan, "CoAtNet: Marrying Convolution and Attention for All Data Sizes," 6 2021. [Online]. Available: https://arxiv.org/abs/2106.04803v2

[27] C. Wang, J. Yang, L. Xie, and J. Yuan, "Kervolutional Neural Networks."

[28] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11211 LNCS, pp. 3–19, 7 2018. [Online]. Available: https://arxiv.org/abs/1807.06521v2

[29] H. Nam, S.-H. Kim, B.-Y. Ko, and Y.-H. Park, "Frequency Dynamic Convolution: Frequency-Adaptive Pattern Recognition for Sound Event Detection." [Online]. Available: https://github.com/frednam93/FDY-SED

[30] H. Nam, S.-H. Kim, and Y.-H. Park, "FilterAugment: An Acoustic Environmental Data Augmentation Method," pp. 4308–4312, 10 2021. [Online]. Available: https://arxiv.org/abs/2110.03282v4

[31] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2019-September, pp. 2613–2617, 4 2019. [Online]. Available: http://arxiv.org/abs/1904.08779 http://dx.doi.org/10.21437/Interspeech.2019-2680

[32] D. Misra, "Mish: A Self Regularized Non-Monotonic Activation Function," 8 2019. [Online]. Available: https://arxiv.org/abs/1908.08681v3

[33] S. Nag and M. Bhattacharyya, "SERF: Towards better training of deep neural networks using log-Softplus ERror activation Function," 8 2021. [Online]. Available: https://arxiv.org/abs/2108.09598v3

[34] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 12 2014. [Online]. Available: https://arxiv.org/abs/1412.6980v9

[35] "Early Stopping Explained — Papers With Code." [Online]. Available: https://paperswithcode.com/method/early-stopping

[36] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," 5 2016. [Online]. Available: https://arxiv.org/abs/1605.06211v1

[37] N. Zeghidour, O. Teboul, F. de Chaumont Quitry, and M. Tagliasacchi, "{LEAF}: A learnable frontend for audio classification," in *International Conference on Learning Representations*, 2021. [Online]. Available: https://openreview.net/forum?id=jM76BCb6F9m

An Efficient Parallel Algorithm for Clustering Big Data based on the Spark Framework

Zineb Dafir*
Faculty of Science of Rabat, Mohammed V University
Rabat, Morocco

Said Slaoui
Faculty of Science of Rabat, Mohammed V University
Rabat, Morocco

Abstract—**The principal objective of this paper is to provide a parallel implementation focused on the main steps of the parameter-free clustering algorithm based on K-means (PFK-means) using the Spark framework and a machine learning-based model to process Big Data. Thus, the process consists of parallelizing the main tasks of the first stage of the PFK-means clustering algorithm using successive RDD functions. Then, the parallel K-means provided by Spark MLlib is invoked by setting the cluster centers and the number of clusters determined in the previous step as input parameters of the parallel K-means. Furthermore, a comparison between the parallel designed algorithm and the parallel K-means was conducted using UCI data sets in terms of the sum of squared errors and the processing time. The experimental results, performed locally using the Spark framework, demonstrate the efficiency of the proposed solution.**

Keywords*—*Clustering; big data; spark; parallel computing; parallel K-means

I. Introduction

Today's digital world is experiencing unprecedented progress, causing the generation of an immense amount of data presented in various formats and derived from many sources, including social networks, the internet of things, mobile apps, multimedia, financial services, and ERP systems. Hence, managing this huge amount of data requires efficient techniques and tools to ensure that the data is processed efficiently to make the right decision according to the application domain [1]. Researchers and scientists are constantly contributing to and developing new machine learning algorithms, as well as designing more efficient frameworks capable of dealing with the continuous flow of data generated each lapse of time [2] [3].

Cluster analysis is one of the appropriate data mining techniques for ensuring efficient Big Data processing [4] [5]. Indeed, the purpose of cluster analysis is to place similar data objects in the same group to construct disjoint clusters. The concept of similarity is firmly based on a distance measure depending on the characteristics of the data object's features. In the same context, several Big Data clustering algorithms were designed based on specific computing platforms aiming to either distribute the clustering tasks on several nodes to perform the necessary calculations in a parallel way or by using a server with high capacities in terms of CPUs, memories, and I/O resources [6] [7] [8].

This paper proposes a new parallel clustering algorithm based on Spark [9], designed to enhance the sequential PFK-means algorithm to leverage distributed systems and process big data. As a result, the proposed parallel implementation is carried out using open-source Hadoop, the Spark framework with RDDs, and a machine learning-based model. More specifically, the proposed algorithm aims to parallelize the two principal stages of the PFK-means algorithm using Spark RDDs functions. The first one consists of applying successive RDDs functions in order to perform the computation of the Euclidean distances, build the initial clusters, and finally update the different clusters to obtain the cluster centers and, consequently, the number k. The second stage invokes the use of the parallel K-means provided by Spark mLlib with as input parameters the cluster centers and the number of clusters discovered in the previous phase. In that respect, a comparison between the developed parallel algorithm and the parallel K-means provided by Spark mLlib was conducted on some UCI data sets based on the sum of the squared error measure and the processing time. The results obtained show that the suggested solution yields more efficient clustering compared to the parallel K-means with the random initialization mode. The main contributions of this paper are as follows:

- Improve the approach to determining the initial cluster centers of the cluster construction step
- Suggest a parallel implementation of the main steps of the PFK-means clustering algorithm based on spark RDDs in order to process big quantitative data.
- Establish a comparison between the designed algorithm and the parallel K-means algorithm provided by spark mLlib.

The remainder of this paper is organized as follows: Section II discusses a literature review. Section III gives a short presentation of parallel computing using the Spark framework. Section IV provides the design and implementation of the parallel developed algorithm. Section V shows the results of the experiment. Finally, Section VI concludes the paper and presents future perspectives.

II. Literature Review

Over the years, clustering algorithms have undergone several evolutions and improvements to face the challenges of managing heterogeneous and complex data generated from multiple and varied digital sources. In particular, K-means [10] [11] is one of the most widely used clustering algorithms due to its simplicity and efficiency in processing data. This algorithm has experienced several adaptations and improvements

*Corresponding authors.

to address initialization issues, exploit distributed systems and handle Big Data [12] [13] [14]. In this context, several research studies based on different parallel computing architectures have been carried out to give birth to powerful and scalable algorithms.

Among these algorithms, a parallel adaptive Canopy-K-means algorithm [15] aims to solve the manual selection problem for the distance threshold T2 in the Canopy process. In this regard, it is improved by adaptive parameter estimation using the MapReduce-based Model and the Spark framework. The clustering results on the Stanford Large Network Dataset Collection (SNAP) data set and self-built Dimension demonstrate the efficiency of the proposed solution.

Another improvement in the parallel K-means is introduced in [16], which is an improved K-Means Clustering Algorithm for Big Data Mining based on the density of data points to construct the corresponding clusters. Thus, the algorithm was parallelized on the Hadoop platform using the distributed database to improve its efficiency and decrease the running time. The simulation results prove that the enhanced solution outperforms the classical K-means and the DBSCAN algorithm in terms of clustering accuracy by 10%.

In the same context, a clustering center selection method [17] was developed to solve the issue of the random nature and limited quality of initial cluster center selection. In addition, a parallel implementation using the Spark computing framework was suggested to perform Big Data clustering and therefore obtain higher execution efficiency, accuracy, and good stability in big data.

More recently, another improved initialization method [18] for the K-means algorithm was implemented to enhance the initial points selection strategy using sparse reconstruction. Moreover, the parallel version of the algorithm was performed based on the MapReduce framework and Hadoop cluster using the real customer data from the JD Mall. The experimental results demonstrate its efficiency in processing large amounts of data.

Another contribution aiming to address the K-means algorithm initialization issue is a parallel clustering algorithm based on grid density [19]. Indeed, the process of the algorithm is based on specific strategies, including, locality sensitive hash function(DP-LSH), the adaptative grouping strategy (AGS), and the MapReduce framework to operate the cluster centers in parallel and therefore increase the performance of the proposed approach.

III. Parallel Computing using Spark Framework

Parallel computing is one of the efficient solutions to process large amounts of data using computing platforms. Indeed, these platforms allow the distribution of the calculations according to different architectures [20]. Such platforms can be either horizontally scaled platforms or vertically scaled platforms [7]. In particular, Spark is a horizontally scalable cluster computing framework [9], performing parallel processing for data-intensive applications with working sets. These applications are managed by the driver program, which allows them to execute the user's main function as well as the parallel operations in a cluster. Hence, it has the advantage of executing tasks by ensuring locality-aware scheduling, fault tolerance, and load balancing [9]. The spark is distinct from MapReduce since it handles iterative jobs and enables users to run queries on a Big Data set by loading into memory only the required data set. More specifically, the Spark framework is structured around two main abstractions: resilient distributed dataset (RDD) and shared variables. The RDD is a collection of items distributed through the cluster nodes that can persist in memory to be reused in the event of parallel operations. The shared variables are also used in parallel operations to share the necessary variables among jobs or between jobs and the driver program. In addition, Spark is powerful and outperforms the Hadoop MapReduce framework by 10× in interactive machine learning workloads. Fig. 1highlights the basics of spark architecture.

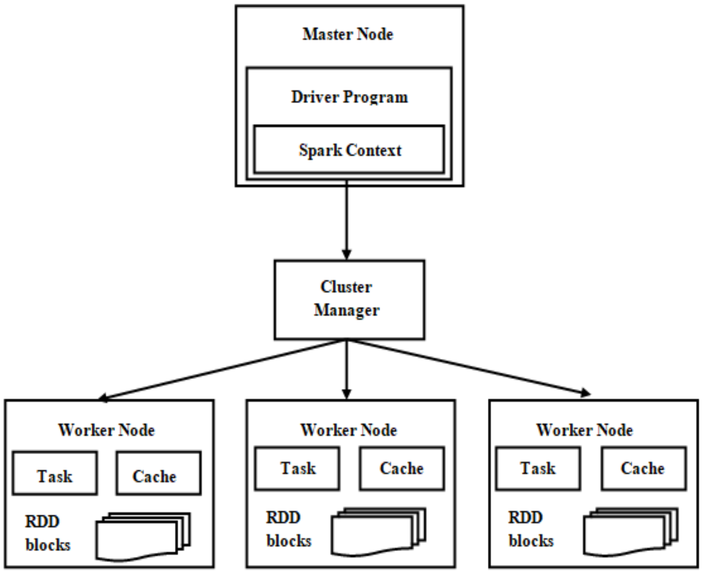


Fig. 1. Spark Architecture.

IV. The Proposed Parallel PFK-means

This section first introduces an overview of the PFK-means clustering algorithm for processing quantitative data, then gives a detailed description of the implementation of the parallel algorithm designed for Big Data processing, and finally presents the algorithm describing the various Spark RDD functions that allow the parallelization of the tasks of the suggested algorithm.

A. The Sequential PFK-means

The PFK-means algorithm [21] is a parameter-free clustering algorithm aiming to construct progressively homogeneous clusters until the appropriate number of clusters is automatically detected. This heuristic is a combination of the E-transitive heuristic [22] adjusted for quantitative data, and the traditional K-means [10] [11]. Indeed, the sequential version of the PFK-means algorithm performs a partitioning clustering based on two main stages. The first one aims to construct clusters successively based on the similarity between the cluster centers and the data objects using the Euclidean

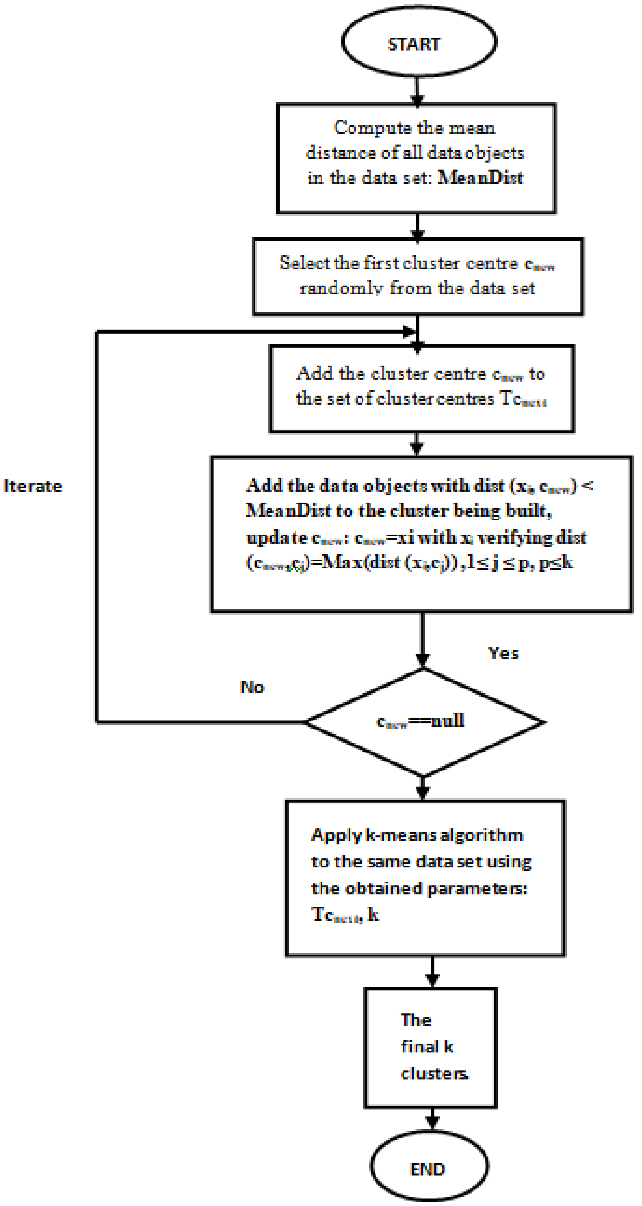

START
Compute the mean distance of all data objects in the data set: MeanDist
Select the first cluster centre c_new randomly from the data set
Add the cluster centre c_new to the set of cluster centres Tc_next
Iterate
Add the data objects with dist (x_i, c_new) < MeanDist to the cluster being built, update c_new: c_new=xi with x_i verifying dist (c_new,c_j)=Max(dist (x_i,c_j)) ,1≤ j ≤ p, p≤k
Yes
No
c_new==null
Apply k-means algorithm to the same data set using the obtained parameters: Tc_next, k
The final k clusters.
END

Fig. 2. Sequential PFK-means Flowchart.

distance calculation and consequently, gets the initial cluster centers and the number of clusters k, which are the two primary input parameters for the K-means algorithm. The second stage consists of applying the K-means algorithm to the same data set taking into consideration the cluster centers obtained and the number of clusters k reached in the first stage. The first stage of the PFK-means algorithm is an independent clustering algorithm that enables the discovery of clusters with appropriate cluster centers. The K-means algorithm is applied to enhance the quality of the clusters obtained. Fig. 2 resumes the main steps of the PFK-means algorithm.

B. Design and Implementation of the Parallel PFK-means

The PFK-means is a sequential and iterative clustering algorithm, suitable for processing unsupervised machine learning data sets. However, the efficiency of this algorithm will be restricted when dealing with a large amount of data. Among the solutions to improve the performance of this algorithm and make it suitable for processing Big Data is the parallelization of its tasks using the open-source Hadoop, Spark RDD, and Machine Learning-based Model. In that respect, the implementation of the parallel version of PFK-means consists of the parallelization of its two main stages. In this way, the first stage consists of the parallelization of three principal tasks, including the calculation of the total average of the Euclidean distances, the construction of the initial clusters, and assigning the data objects to the appropriate clusters as well as updating the discovered cluster centers. The next stage involves the use of the parallel K-means provided by spark mLlib, using as input parameters the discovered cluster centers as well as the number of clusters automatically detected in the first stage. Fig. 3 illustrates the design of the parallel implementation of the PFK-means. In the following, the steps of the initial stage will be thoroughly described.

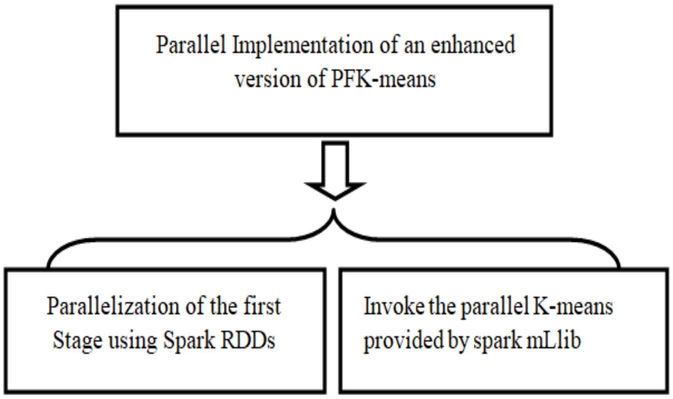

Parallel Implementation of an enhanced version of PFK-means
Parallelization of the first Stage using Spark RDDs
Invoke the parallel K-means provided by spark mLlib

Fig. 3. The Design of the Parallel Implementation.

1) Calculation of the Total Average of the Euclidean distances: In order to compute the total average of the Euclidean distances, the process starts by reading the data set being processed locally into RDDs partitions since it is the first step. Then these RDDs' partitions are transformed using the Map function with the purpose of calculating the Euclidean distance between each two data objects, which is performed in several iterations. Therefore, in each iteration, the Reduce function sums the partial Euclidean distances for each partition Map. Then, after completing all the iterations, another Reduce function is applied to the final result, in order to obtain the total average of the Euclidean distances. Fig. 4 depicts the process of spark RDDs to compute the total average Euclidean distance of the processed data set.

Although the use of Spark is optimal since it avoids storing the data on a hard disk, the calculation of the average of the Euclidean distances of the data set is costly. It is, therefore, essential to optimize this calculation. The first way involves calculating the average of the Euclidean distances of a sample instead of processing the entire data set using the Map and Reduce functions. The second solution consists of first reordering the whole data set randomly and then applying the MapPartitions function, which allows decomposing the data

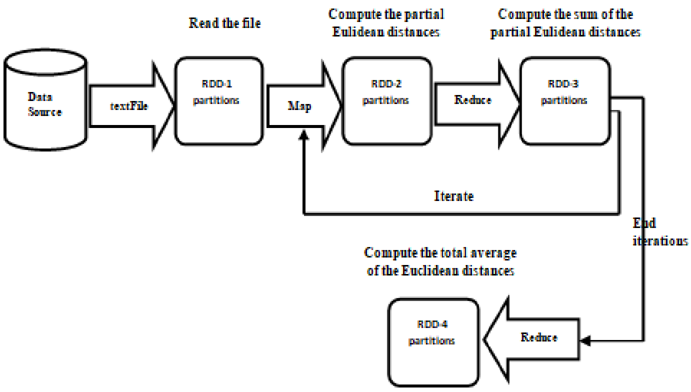

Read the file
Compute the partial Eulidean distances
Compute the sum of the partial Eulidean distances
Data Source
textFile
RDD-1 partitions
Map
RDD-2 partitions
Reduce
RDD-3 partitions
Iterate
End iterations
Compute the total average of the Euclidean distances
RDD-4 partitions
Reduce

Fig. 4. The Parallel Process of the Average of Euclidean Distance Calculation.

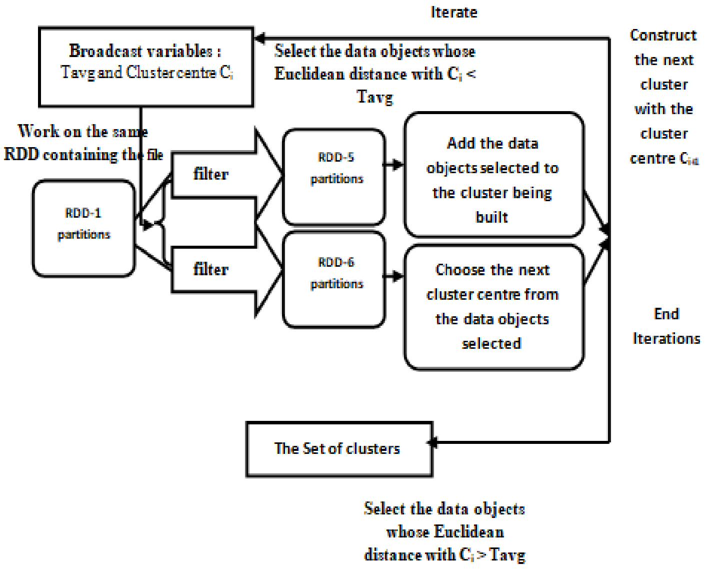

Iterate
Broadcast variables : Tavg and Cluster centre Ci
Select the data objects whose Euclidean distance with Ci < Tavg
Construct the next cluster with the cluster centre Ci+1
Work on the same RDD containing the file
filter
RDD-5 partitions
Add the data objects selected to the cluster being built
RDD-1 partitions
filter
RDD-6 partitions
Choose the next cluster centre from the data objects selected
End Iterations
The Set of clusters
Select the data objects whose Euclidean distance with Ci > Tavg

Fig. 5. The Parallel Process of Initial Cluster Construction.

set into a specific number of partitions. Then, each partition performs the computation of the two-by-two Euclidean distances of a part of the data. Thus, the result produced by the MapPartitions function is transmitted to the Reducer, allowing the calculation of the total average of the Euclidean distances. In this solution, the MapPartitions and Reduce functions are invoked only once, so the calculation time is much reduced compared to the solution explained previously.

2) The Construction of the Initial Clusters: This step starts by choosing the first data object in the data set as the first cluster center. Then, the data objects similar to this cluster center are gathered in such a way that the Euclidean distance between this cluster center and each data object in the whole data set is less than the total average of the Euclidean distances calculated in the previous step. Contrary to the sequential version of the algorithm in which the construction of the first cluster is realized in n iterations, the construction of the first cluster in the parallel version of the algorithm is done using the filter function, which selects the elements similar to the cluster center by calculating the Euclidean distances in parallel. Subsequently, the cluster center of the second cluster is determined in such a way that the data object similar to the first cluster center whose Euclidean distance is the greatest is compared to the data objects not similar to the center, so the data object whose Euclidean distance yields an average value is chosen as the second cluster center. Consequently, the determination of the cluster center is realized using the filter function, which allows selecting the cluster center among the objects not similar to the first cluster center by calculating the Euclidean distances in parallel. In this way, the other cluster centers are determined successively to follow the same process. This process is finished when all the data objects are classified and there are no more data objects to choose from as a new cluster center. And lastly, it is important to mention that during the construction of the clusters, each data object can be placed in several clusters, and therefore the clusters achieved at the end of this step are overlapped. This allows for the enlargement of the size of the clusters and, accordingly, improves the quality of the cluster centers obtained. Fig. 5 shows the cluster construction process using the Spark RDDs functions.

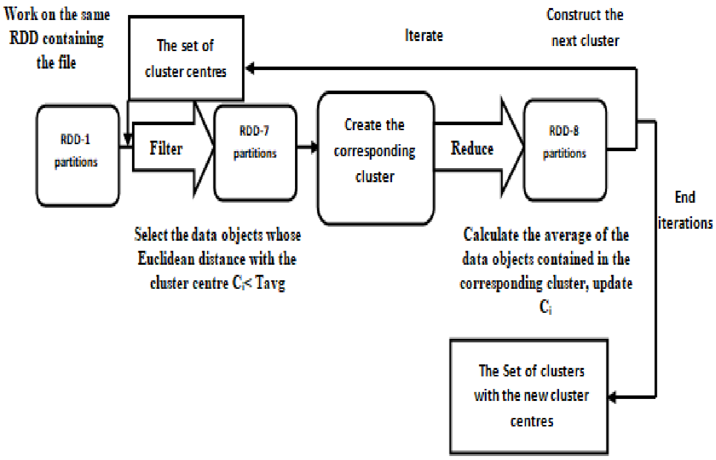

Work on the same RDD containing the file
The set of cluster centres
Iterate
Construct the next cluster
RDD-1 partitions
Filter
RDD-7 partitions
Create the corresponding cluster
Reduce
RDD-8 partitions
End iterations
Select the data objects whose Euclidean distance with the cluster centre Ci< Tavg
Calculate the average of the data objects contained in the corresponding cluster, update Ci
The Set of clusters with the new cluster centres

Fig. 6. The Parallel Process of Updating Clusters.

3) Assigning the Data Objects to the Appropriate Clusters and Update Cluster Centers: After having built the initial clusters in the previous step, this step consists of assigning to each cluster center found in the previous step the data objects similar to it by using the filter function. This function selects the data objects by performing the calculations of the Euclidean distances in parallel. Then, the Reduce function is applied to the RDD returned by the Filter function to update the value of the cluster center by computing the average of the data objects contained in the corresponding cluster. In this step, each data object can be associated with only one cluster, and consequently, there will be no overlaps in the resultant clusters. Fig. 6 describes the process of obtaining the final hard clusters with the new cluster centers using the Filter and Reduce functions. Once the hard clusters with updated cluster centers and the number of cluster centers detected automatically have been obtained, the next step is to apply the parallel K-means provided by Spark mLlib with the input parameters obtained from the first stage, which are the cluster centers and the number of clusters.

C. The Parallel Proposed Algorithm

The following pseudo-code (Algorithm 1) describes the stages of the parallel implementation of the PFK-means algorithm. Hence, the process begins by reading the file containing the set of quantitative data (lines 1 and 2). Subsequently, it seeks to calculate the average of the Euclidean distances of the data set using one of the previously explained methods (line 3). The next step is to build the initial clusters to obtain the list of cluster centers to be used in the next step (from 4 to 17). Afterward, the process aims to assign the data objects to the appropriate centers and update the different centers (from 18 to 24). The last step intends to invoke the parallel K-means provided by Spark mLlib using the parameters obtained from the previous step (from 25 to 27).

Algorithm 1 The Parallel Implementation of PFK-means

```
Input:A set of $n$ data objects $X$
Output:The initial cluster centers $Tc_{next}$. The number of clusters automatically computed
begin
 1: ReadFile = sc.textFile(”the file path”)
 2: dtSet = ReadFile.map(lines).cache()
 3: TotalAverage = ComputeTotalAverage()
 4: sc.broadcast($TotalAverage$)
 5: $NextCenter = newList[0]$, $Next = True$
 6: while $Next$ do
 7:     sc.broadcast($NextCenter$)
 8:     Cluster = dtSet.filter($dist(X, NextCenter) < TotalAverage$)
 9:     NextC= dtSet.filter($dist(X, NextCenter) > TotalAverage$)
10:     SortedList = NextC.sortBy($dist(X, FarthestElement)$)
11:     $NextCenter = MeanValue(SortedList.collect())$
12:     Add the constructed Cluster to $Tc_{next}$
13:     if $NextCenter == Null$ then
14:         $Next == False$
15:     end if
16: end while
17: Save the set of centers in $Centers$
18: $i \leftarrow 0$
19: while $i < ClustersSize$ do
20:     Cluster = dtSet.filter($dist(X, Centers[i]) < TotalAverage$)
21:     UpdateCenter=Cluster.reduce($computeMean(X, Y)$)
22:     Remove the filtred elements from the dataset dtSet
23:     Save the new cluster center in $NewCenters$
24: end while
25: $t_{max} = maxIterations$,$dt = dtSet$,$k = NewCentersSize$
26: $initializationMode = KMeansModel(NewCenters)$
27: FinalC=KMeans.train($dtSet$,$k$,$t_{max}$,$initializationMode$)
end begin
```

TABLE I. Description of the Real-life Data Sets

data set	Data size	Attributes	Cluster number
Iris	150	4	3
Wine	178	13	3
Pima Indian Diabetes	768	8	2
Letter-Recognition(LR)	20000	16	26

V. Experiments

For the purpose of implementing the parallel version of PFK-means with Spark RDDs in stand-alone mode, a machine of 12GB memory and 1TB hard disk was configured with Spark (version 3.0.3) and Hadoop (2.7 version) and PyCharm IDE by installing useful Pyspark libraries (PySpark version is 3.9). Subsequently, the proposed algorithm was evaluated on real-world data sets extracted from the UCI machine learning repository based on the sum of squared errors and the execution time to measure the clustering performance. In addition, the parallel PFK-means were compared with the parallel K-means provided by the mLlib library to demonstrate the efficiency of the developed solution. It is also crucial to notice that the parallel implementation of PFK-means has been tested locally in the PyCharm IDE and, therefore, the size of the data sets has been limited.

A. Data Sets Description

Table I provides a short description of three real-world datasets, extracted from the UCI machine learning repository [23], serving to validate the efficiency of the parallel implementation of PFK-means, including Iris, Wine, and Letter Recognition (LR). In this way, each data set is represented by a data size, a determined number of clusters, and the number of attributes that constitutes the vectors of the data sets.

B. Results Achieved in Terms of the Sum of Squared Errors

One of the effective metrics to evaluate the quality of clustering is the sum of the squared errors. This measure is obtained by calculating the sum of the Euclidean distances between each cluster center and the set of data objects belonging to this cluster. Therefore, the sum of the results found corresponds to the sum of squared errors of the set of clusters. Accordingly, the more minimal this measure is, the better the result is achieved.

Table II reports the sum of squared errors of the clustering result obtained by running the parallel PFK-means and the parallel K-means provided by Spark mLlib on the Letter Recognition (LR) data set. The execution of parallel PFK-means performs clustering by automatically detecting the number of clusters that can be varied by changing the position of the new cluster center determined in each iteration of the cluster construction step of the algorithm. Thus, the reported results are obtained by iterating the parallel K-means one, three, and ten times, respectively. In other words, for the parallel PFK-means, the reiteration is applied just to the second phase of the algorithm, which consists of applying the parallel K-means implementation of Spark mLlib using the cluster centers and the number of clusters found in the first phase of the algorithm. In this respect, the acquired results are compared

TABLE II. The Sum of Squared Errors of the Clustering Results on Letter-Recognition Data Set

number of clusters	parallel PFK-means	parallel K-means (mLlib)	number of iterations
24	**117015.22**	118400.60	1
24	**113035.25**	113995.74	3
24	111162.61	**110717.19**	10
26	**114813.58**	118262.82	1
26	**111198.35**	112354.80	3
26	**109506.13**	109853.67	10

TABLE III. The Sum of Squared Errors of the Clustering Results on Iris, Wine, and Pima Data Sets

data set	parallel PFK-means	parallel K-means (mLlib)
Iris	**97.34**	97.70
Wine	**18889.14**	19015.23
Pima	**49403.81**	64315.08

with the parallel K-means of Spark mLlib using the random initialization mode.

By observing the results displayed in the Table II, it can be seen that the parallel PFK-means outperforms the parallel K-means in terms of the sum of squared errors, except for the tenth iteration with k = 24, for which the parallel K-means exceeds the proposed algorithm. It is also noteworthy that the PFK-means produces good results from the first iteration, implying that the cluster centers discovered in the first phase of the algorithm are well-positioned in comparison to the cluster centers discovered using the random initialization mode. Besides the fact that the parallel PFK-means allows discovering the number of clusters automatically, the result given by the developed algorithm is stable. Table III presents a comparison of the parallel PFK-means and the parallel K-means in terms of the sum of squared errors for the real-world data sets Iris, Wine, and Pima. The results clearly show that the execution of the parallel K-means after determining the input parameters (number of clusters and the cluster centers) from the first phase of the developed parallel algorithm consumes less time compared to the execution of parallel K-means with the random initialization mode for the three data sets used.

C. Results of the Processing Time

In order to evaluate the performance of the parallel PFK-means concerning running time, the second phase of the proposed algorithm has been compared with the parallel K-means of Spark mLlib using the random initialization mode, which allows to randomly generate the cluster centers. In this case, the number of clusters k used is the number detected by the first phase of the parallel PFK-means, and therefore the number of cluster centers must be fixed before the execution. On the other hand, by running the second phase of the PFK-means, the number of clusters is automatically detected and assigned to the parallel K-means as well as the cluster centers.

Fig. 7 illustrates the execution time of the second stage of the parallel PFK-means compared to the execution time of the parallel K-means on Iris, Wine, Pima, and Letter-Recognition data sets. Fig. 7 shows that the PFK-means outperforms the parallel K-means for all the data sets used.

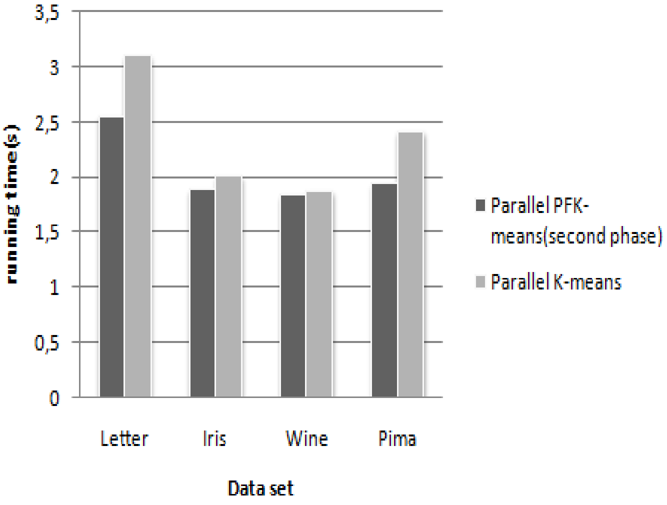


Fig. 7. Clustering Time on Real-world Data Sets.

D. Discussion of the Obtained Results

According to the experiments that were conducted on real-world data sets, it was proved that the proposed algorithm is a parameter-free clustering algorithm since it can automatically determine the set of cluster centers and the number of clusters without specifying any input parameters, and therefore the suggested solution is suitable for unsupervised learning. However, running the parallel version of k-means provided by Spark mLlib requires fixing the number of clusters as well as specifying the initial cluster centers. Moreover, the execution of the developed algorithm using real-world data sets gives significant results compared to the results achieved by the parallel implementation of K-means provided by Spark mLlib in terms of the sum of squared errors and execution time. In addition to that, the parallel implementation of the method allows the distribution of the tasks constituting the main steps of the algorithm on several nodes, which allows for the processing of very large files.

VI. Conclusion and Perspective

The main concern of this paper is to design a new parallel clustering algorithm based on the main steps of the PFK-means clustering algorithm for processing large quantitative data sets by bringing a significant improvement in the way of determining the initial cluster centers in the cluster construction step. Thus, it exposes a general presentation of the parallel computing platforms and the main process of the Spark framework. Moreover, this paper presents a preview of the sequential version of the PFK-means algorithm [21]as well as a detailed explanation of the proposed parallel algorithm. Furthermore, experiments based on UCI data sets were conducted based on the sum of squared errors and the execution time. The results have clearly shown that the suggested parallel algorithm is an efficient clustering algorithm as it can automatically construct the initial cluster centers as well as the number of clusters without having to specify any parameters beforehand. Besides, it has been demonstrated that the developed algorithm outperforms the parallel K-means provided by Spark mLlib in terms of squared errors and processing time.

In our future research, we intend to concentrate on creating a remote server configuration to distribute the different tasks of the algorithm on multiple nodes and therefore process very large files. In addition, we will establish a deep analysis to compare both the stand-alone mode implementation and cluster mode implementation. Furthermore, we will focus on the implementation of the proposed algorithm using other similarity measures and various data sets. It may also be possible to apply the proposed solution to perform real-time processing.

References

[1] M. Naeem, T. Jamal, J. Diaz-Martinez, S. A. Butt, N. Montesano, M. I. Tariq, E. De-la Hoz-Franco, and E. De-La-Hoz-Valdiris, "Trends and future perspective challenges in big data," in *Advances in intelligent data analysis and applications*. Springer, 2022, pp. 309–325.

[2] A. E. Ezugwu, A. M. Ikotun, O. O. Oyelade, L. Abualigah, J. O. Agushaka, C. I. Eke, and A. A. Akinyelu, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110, p. 104743, 2022.

[3] K. El Handri and A. Idrissi, "Parallelization of $top\_\{k\}$ algorithm through a new hybrid recommendation system for big data in spark cloud computing framework," *IEEE Systems Journal*, vol. 15, no. 4, pp. 4876–4886, 2020.

[4] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[5] R. Garcia-Dias, S. Vieira, W. H. L. Pinaya, and A. Mechelli, "Clustering analysis," in *machine learning*. Elsevier, 2020, pp. 227–247.

[6] H. Ibrahim Hayatu, A. Mohammed, and A. Barroon Isma'eel, "Big data clustering techniques: Recent advances and survey," *Machine Learning and Data Mining for Emerging Trend in Cyber Dynamics*, pp. 57–79, 2021.

[7] Z. Dafir, Y. Lamari, and S. C. Slaoui, "A survey on parallel clustering algorithms for big data," *Artificial Intelligence Review*, pp. 1–33, 2020.

[8] Y. Lamari and S. C. Slaoui, "Pdc-transitive: An enhanced heuristic for document clustering based on relational analysis approach and iterative mapreduce," *Journal of Information & Knowledge Management*, vol. 17, no. 02, p. 1850021, 2018.

[9] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets," in *2nd USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 10)*, 2010.

[10] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.

[11] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[12] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020.

[13] A. Ashabi, S. B. Sahibuddin, and M. Salkhordeh Haghighi, "The systematic review of k-means clustering algorithm," in *2020 The 9th International Conference on Networks, Communication and Computing*, 2020, pp. 13–18.

[14] S. Bouhout, Y. Oubenaalla, and E. H. Nfaoui, "Comparative study of two parallel algorithm k-means and dbscan clustering on spark platform," in *International Conference on Advanced Intelligent Systems for Sustainable Development*. Springer, 2020, pp. 245–262.

[15] D. Xia, F. Ning, and W. He, "Research on parallel adaptive canopy-k-means clustering algorithm for big data mining based on cloud platform," *Journal of Grid Computing*, vol. 18, no. 2, pp. 263–273, 2020.

[16] W. Lu, "Improved k-means clustering algorithm for big data mining under hadoop parallel framework," *Journal of Grid Computing*, vol. 18, no. 2, pp. 239–250, 2020.

[17] X. Lu, H. Lu, J. Yuan, and X. Wang, "An improved k-means distributed clustering algorithm based on spark parallel computing framework," in *Journal of Physics: Conference Series*, vol. 1616, no. 1. IOP Publishing, 2020, p. 012065.

[18] Y. Liu, X. Du, and S. Ma, "Innovative study on clustering center and distance measurement of k-means algorithm: mapreduce efficient parallel algorithm based on user data of jd mall," *Electronic Commerce Research*, pp. 1–31, 2021.

[19] Y. Mao, D. Gan, D. S. Mwakapesa, Y. A. Nanehkaran, T. Tao, and X. Huang, "A mapreduce-based k-means clustering algorithm," *The Journal of Supercomputing*, vol. 78, no. 4, pp. 5181–5202, 2022.

[20] B. Balusamy, S. Kadry, A. H. Gandomi *et al.*, *Big Data: Concepts, Technology, and Architecture*. John Wiley & Sons, 2021.

[21] S. Slaoui and Z. Dafir, "A parameter-free clustering algorithm based k-means," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 3, 2021.

[22] S. C. Slaoui, Z. Dafir, and Y. Lamari, "E-transitive: an enhanced version of the transitive heuristic for clustering categorical data," *Procedia Computer Science*, vol. 127, pp. 26–34, 2018.

[23] A. Asuncion and D. Newman, "Uci machine learning repository," 2007.

Development of a Low-Cost Teleoperated Explorer Robot (TXRob)

Rafael Verano M., Jose Caceres S., Abel Arenas H., Andres Montoya A.,
Joseph Guevara M., Jarelh Galdos B., Jesus Talavera S.
Universidad Nacional de San Agustín
de Arequipa, Perú

***Abstract*—Natural disasters such as earthquakes or mudslides destroy everything in their path, causing buildings to collapse, which can cause people to lose their lives or suffer permanent injuries. Rescuers and firefighters are responsible for entering these ruined buildings, this work being very dangerous for them because they can get trapped in the rubble or suffocate due to the harmful gases found inside these buildings. Taking into consideration the risk in this type of operations, technological innovations can be used to help in the exploration of ruined buildings and the rescue of people. Therefore, this article describes the development of TXRob, a low-cost teleoperated robot used in the exploration of post-disaster scenarios. TXRob has artificial vision, environmental gas recognition sensors, a real-time data display panel, is sized to enter buildings, and is capable of moving over uneven surfaces, such as debris or cracks, thanks to its track system. A human operator can remotely monitor and control the robot. The TXRob's versatility as well as sensors performance has been tested on uneven and harsh surfaces in a simulated disaster environment. These tests suggest that the designed robot is suitable for use in rescue situations.**

***Keywords*—Rescue robot; teleoperation; low cost robot; artificial vision**

I. Introduction

The Peruvian territory has suffered serious seismic episodes due to its location in the fire ring [1], [2]. Some of these earthquakes have occurred on June 23, 2001 with a magnitude of 8.4Mw (Moment scale), this event happened in Arequipa, leaving more than 240 dead, 70 people missing and 2400 injured [3]; in the province of Pisco on August 15, 2007, a 7.9Mw earthquake occurred, leaving 95 dead, 2,291 injured, 76,000 houses destroyed and uninhabitable and 431,000 people affected [4]; on May 26, 2019, an earthquake of 8.0Mw occurred in Alto Amazonas, which caused the death of 2 people and several injured people [5]. Due to these disasters, contingency procedures were developed to minimize the damage caused. In addition, systems capable of predicting these telluric movements have been created in order to minimize the post-disaster impact [6], but one must indeed have sufficient capacity to respond to a natural disaster, not only an earthquake, but also landslides or fires where the structure of the building of the place is affected.

After a disaster, access to the building is partially or totally inaccessible, so a teleoperated robot would assist with contingency programs in these hazardous situations, but the robot must be able to overcome obstacles and accurately locate potential survivors. For example, in the article [7] a rescue robot for miners trapped in a cave-in, equipped with cameras and sensors that detect toxic gas levels. Several robotics projects robotic projects have been designed with this approach, most of them remotely controlled to support human operators in SAR (search and rescue operations) tasks [8], [9].These robots work more efficiently and faster than a human human agent, which is reflected in the higher number of rescued people, this is the main objective of the robot. Rescue tasks are critical and are performed in dangerous environments for people, further reinforcing the need to employ this type of robots [10].

There are several approaches when designing a rescue robot. To understand and analyze the post-disaster environment, unmanned aerial vehicles (UAV) or unmanned ground vehicles (UGV) are used, an example of these vehicles is the Tactical Hazardous Operations Robot (THOR) [11]. In order to map a post-disaster scenario, SLAM (Visual Simultaneous Localization and Mapping) technology is used [12]. When rescuing survivors, the impact of the rescue robot on the survivor must be taken into account. Parameters such as the force induced by the robot and the speed of translation are important to analyze [13]. But these methods prove to be costly [14], this is why the development of a low-cost robot would be more accessible to the rescue teams.

In 2019, air pollution was considered the greatest environmental health risk. environmental health risk. The main sources of air pollution are large-scale burning of wood, biomass or crop residues, fuel adulteration, uncontrolled emissions from vehicles and factories, traffic congestion and accelerated construction. These sources cause smog and thus increase airborne particulate matter (PM_10, $PM_{2,5}$), NO_x, NH_3, SO_x , CO and other VOC (volatile organic compounds) in the air [15]. It is very likely that the concentration of gases after a disaster is not optimal for humans; this is due to gas leaks, fires caused by short circuits or ruptures of pipes inside buildings; there are several sensors that can detect these measurements but it is necessary to take these measurements remotely. Normally a gas sensor is incorporated to determine these parameters and thus secure the area for human intervention. These sensors are: CO_2 sensor, VOC sensor (Volatile Organic Compounds Sensor), NO_X sensor, PIR sensor, among others [16], [17].

Good body temperature control is vitally important in a survivor, because monitoring body temperature can predict stroke; sensors such as the UHF RFID temperature sensor are used to obtain a more accurate measurement [18], [19]. Estimating the maximum response time that can be expected is vital to prevent the survivor from suffering severe respiratory distress due to hypothermia, considered below 35C°, which could result in death [20]. There are several ways to detect life signs in survivors in a post-disaster scenario, considering that there is a 24-hour life time frame, these signs become a priority for the rescue robot operator. Sound acquisition serves as a possible support to other sensors in order to locate survivors quickly and efficiently [21], [22]. In today's technology, the use of infrared cameras is necessary for the detection of heat signatures within an environment to be explored. Due to the relevance of this tool in the exploration of environments for the detection of people, the processing of images from the camera is performed [23], [24].

The stereo vision has the advantage of being wireless, so it is the main component of the artificial vision modules of the detection systems in robots [25]. An application is the tracking of objects, where it is necessary to estimate the distance of the object, thus achieving a high accuracy in real time [26]. Being complementary to other systems such as ”leap motion” for interfaces ”man-robot” for applications in robots EOD (bomb disposal) [27].

An anti-shock system is used to assist the operator in manipulating the robot by avoiding obstacles that may pass unnoticed by the operator. The main components of this system are usually infrared sensors, ultrasonic sensors, among others [28].

In the communication with the robot and the control center there are several options, for example the Zigbee model which has a central coordinator, routers and end devices using as XBee-PRO ZB S2B module [29], but taking into account the low cost objective of the project, it is considered to create a local WiFi network configured with the Raspberry Pi 4 controller. Due to studies conducted on different platforms of the different versions of Raspberry, it is concluded that the latest versions support better the WiFi network [30].

This document is divided as follows: The Methodology is presented in Section II. Section III describes the development of TXRob. Section IV presents the development of the control center. Section V presents the results obtained. Finally, Section VI presents the conclusions of the research.

II. Methodology

This paper presents the development of an exploration robot for post-disaster areas, TXRob. The reason for the development of the TXRob robot is the great need of agents dedicated to the rescue of people trapped in ruins caused by a natural or human disaster. The main features of TXRob are:

- Ability to overcome rugged surfaces and debris, dust and moisture resistance, and teleoperated control.
- Perfect for simple fabrication due to its low cost.
- Weight less than 10Kg and dimensions of 20cm x 30cm x 15cm, according to the category of exploration robots.
- It has gas, audio and webcam sensors to collect data from the environment where it is located and transmit them to the operator through its graphic interface located in the control station.

Fig. 1 is the representation of the complete system in a block diagram. First, the control station generates a wifi server, the TXRob robot connects to this network as a client. Through this network, the robot sends the data from the sensors and the video captured in real time. The values obtained by the gas sensors will serve to determine the environmental conditions, while the video, the microphone and the ultrasound will allow the detection of people trapped in the rubble. Later, again at the control station, the operator visualizes this information on the user interface and through a command can generate instructions that control the movement of TXRob through the motor drivers. A bidirectional communication is generated between the robot and the control station.

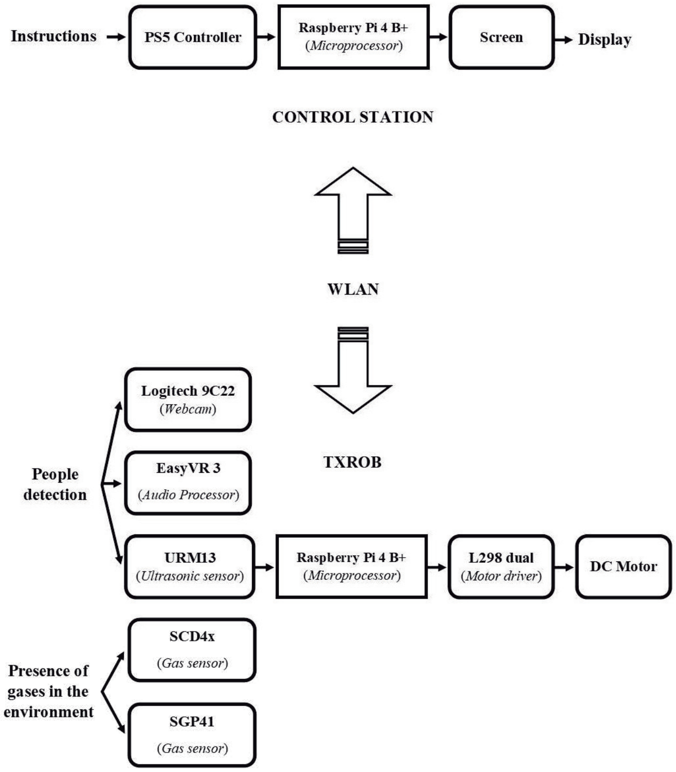


Fig. 1. System Block Diagram

The performance of the robot was measured in two evaluations, the first was an evaluation of the sensors in two different situations: an environment under normal conditions and an environment simulated under disaster conditions using a gradual and sustained increase in temperature, humidity and CO2. The second evaluation was focused on the versatility in displacement of TXRob on a flat surface and a rough surface, where the time needed to reach a given distance in meters was obtained.

III. TXRob Development

A. Mechanical Structure Design

SolidWorks software was used to produce the 3D design and generate the 2D drawings. The chassis of the robot is designed to be made of stainless steel, for its high durability and strength, so that the electronic components are fully protected. The motion system has a hybrid configuration between the rocket-buggie system, used in Martian rover explorers, and a crawler system, see Fig. 2 and 3. This rear-tracked configuration takes advantage of both systems; on the one hand, the left and right rocker suspensions are connected to the body through a differential balancing mechanism, which decreases chassis instability and pitching, while the tracked system is ideal for exploring rough terrain. The two front wheels allow for better handling and more efficient turns. Ultrasonic sensors are mounted front and rear to avoid collisions with the environment, and the top-mounted camera allows a greater field of vision.

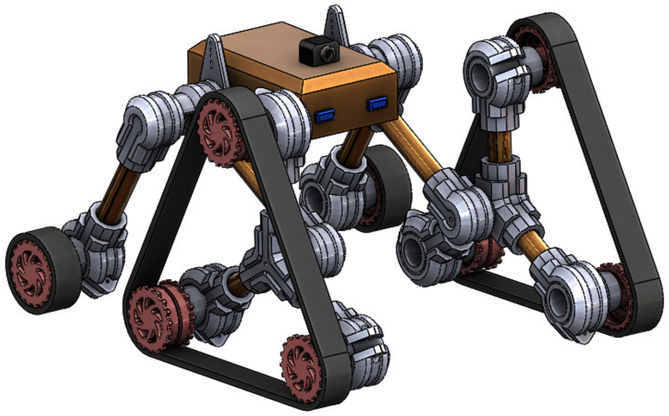


Fig. 2. Isometric View of the Mechanical Structure Design.

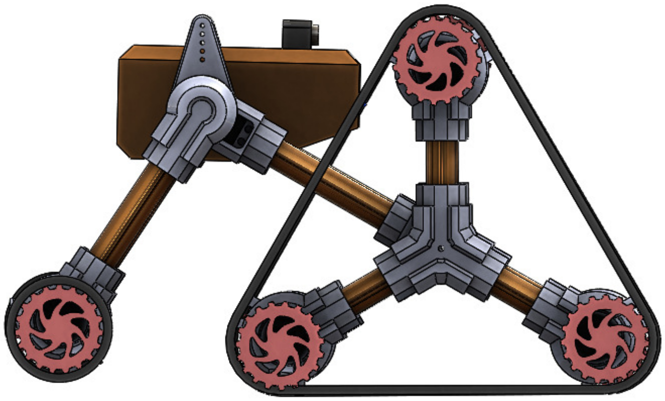


Fig. 3. Side View of the Mechanical Structure Design.

B. Electronic Circuit Assembly

For the development of the electronic and schematic design, the EasyEDA software with free license is used, available at the URL https://easyeda.com/es. This software has proven to have good performance and a fast learning curve at the basic and intermediate levels. Fig. 4 shows all the connections made for TXRob in addition to proper pin labeling. The component connections of the TXRob robot is centered on the Raspberry Pi 4 model B+ board, this Raspberry Pi model is the most recent on the market and has 8GB (gigabytes) of RAM, necessary to perform the video processing, data transmission and motion control of TXRob. This Raspberry Pi board is powered by a 7805 voltage regulator, shown at the beginning of the electronic diagram in Fig. 4, through pin 2, in addition this 5v power pin is also power supply for the sensors and webcam, and also provides a control level for the motor drivers (L298 dual). The L298 driver in addition to using 5v also requires a 12v source to power the motors, this voltage is supplied by the batteries through the VIN pin. The VIN pin is also the input to the 7805 regulator circuit.

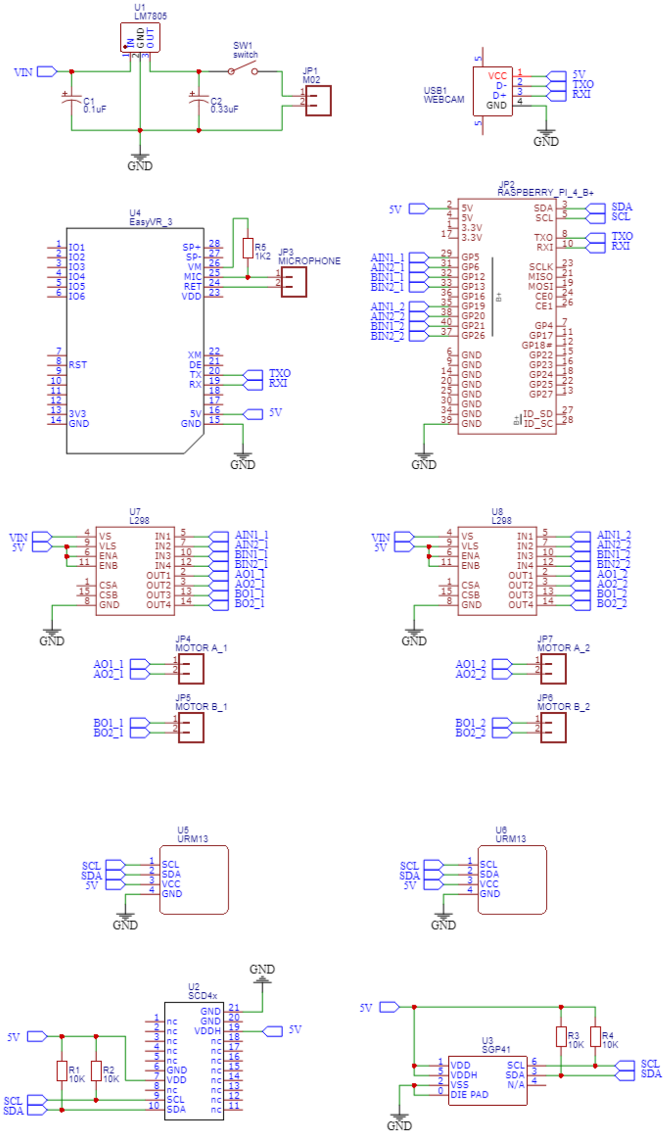


Fig. 4. Schematic Diagram.

1) Gas Sensors: In the exploration stage after a disaster, the environments are more closed, which increases the concentration of gases harmful to humans, so the measurement of these gases is a relevant point for rescuers.

- CO_2 Sensor: The CO_2 is not really a toxic gas but it produces the displacement of oxygen and in high concentrations of more than 30,000 ppm, it can produce asphyxiation. The use of this sensor will allow us to measure the ppm of the explored areas and to be able to analyze the different levels (See Table I). To determine the ppm values we have used the SCD41 sensor that allows us to measure a wider range, between 400 and 5000 ppm, has a typical minimum consumption of only 0.45mA at 3.3V and 0.36mA at 5V, which allows us to create CO2 meters with great autonomy.

TABLE I. CO_2 LEVELS

CO_2 **levels.**	**Air Quality**
250 -450 ppm	Typical atmospheric concentration
450-600 ppm	Acceptable concentration in the interior.
600-1000 ppm	Concentracion aceptable en el interior
1000-2500 ppm	Tolerable concentration. Drowsiness
2500-5000 ppm	Tolerable concentration. Headache. Stagnant air. Loss of concentration and attention.
6000-30000 ppm	Dangerous concentration, only light exposures.
30000-50000 ppm	Extreme concentration. Intoxicant
$>$ 50000 ppm	Extreme concentration. Permanent brain damage and death.

- Volatile Organic Compounds Sensor (VOC): Volatile Organic Compounds are air pollutants that, when mixed with nitrogen oxides, react to form ozone. The presence of high concentrations of ozone in the air we breath is very dangerous. The measurement of $VOCs$ is also relevant to the air quality in a space.
- NO_X Sensor: NO_2 is a toxic and irritating gas. Continued exposure to NO_2 is associated with various respiratory tract diseases.

For the measurement of VOC and NO_X, the SGP41 sensor was used, which has a measurement range from 0 to 1000ppm of ethanol equivalents and 0 to 100ppm of NO_2 equivalents, with a response time of less than 10s for VOC and less than 250s for NO_X necessary for real time sampling to determine the air quality of the area to be examined.

2) Ambiental Sensors:

- Temperature Sensor: The temperature in an environment is an essential indicator that can give us data, the presence of heat can give us to understand the presence of a person who is stuck in the debris.
- Humidity Sensor: Humidity is a factor that we must take into account, the excess of it in an environment can generate diseases or respiratory conditions.

3) Proximity Sensors: To support the teleoperation of the robot, it was decided to use these sensors, which allow us to measure distances and object detection. For the design, they were used for object detection and for an anti-shock system of the robotic structure.

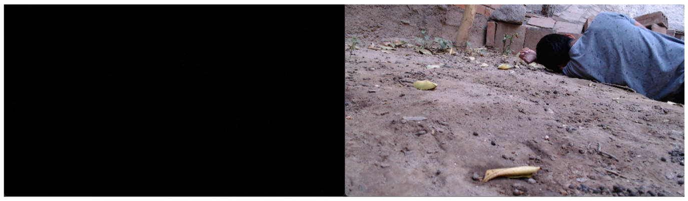
- Infrared sensor: The infrared sensor is used in the market to measure distances in environments with absence of light which will allow us to detect obstacles within the area to be explored.
- Ultrasonic Sensor: The use of ultrasonic sensor as a support in the measurement of distances will help us to obtain more reliable values for the manipulation of the robot. The URM13 ultrasonic sensor has an excellent sensitivity with an effective measurement range of 15 cm to 900 cm with 1% accuracy and I2C communication.

4) DC Motor with Caterpillar: Greartisan DC motors are ideal for the movement of the robot. The proposed crawler system allows the robot to be transported over rough terrain and the powerful motors allow this system to be put into operation. The working voltage of the motors is 12v DC. The gearbox ratio is 1:22, with the specifications being: rated torque of 2.2Kg.cm, rotational speed of 200RPM and rated current of 100mA.

5) Camera: Vision is relevant in scanning robots in order to provide visual information of the operation environment. In addition, a camera was implemented to provide real-time video of the scanned scenario. Due to the costs, a USB camera was chosen. To achieve a teleoperation, through programming in Python language and the Flask framework, the real-time video display was converted to the WiFi network generated by the RaspBerry of the control station, i.e., the USB camera was configured to work as an IP camera, thus saving costs and obtaining the same benefits.

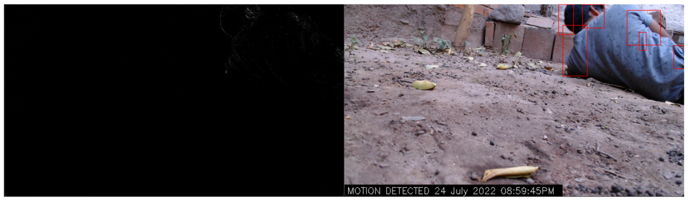
C. Motion Detection Algorithm

The developed detection algorithm performs a subtraction between a first frame captured by the camera ($t = n$) and the next frame ($t = n + 1$), when the value of each pixel obtained by this difference exceeds a threshold , considered threshold of 100 for presenting better results, the algorithm sends an alert to the operator, indicating that the robot has detected a person. Fig. 5 shows a person lying on the ground, which does not move, for that reason, the algorithm does not place red rectangles in the right window, and the left window is completely black. On the other hand, in Fig. 6, white areas are shown on the left, this is because the person has stood up, in addition, on the right there are six red rectangles around the person.



a) b)

Fig. 5. Camera Frame without Motion Detection. a) Differential Window. b) Original Video Window Plus Motion Limiting Zones.



a) b)

Fig. 6. Camera Frame with Detected Motion. a) Differential Window. b) Original Video Window Plus Motion Limiting Zones.

IV. Control Station Development

A. Robot Control

The robot is controlled by a PS5 Dualshock5 controller connected to a Raspberry Pi (Control Station), which through wireless communication sends the instructions received by the controller to another Raspberry Pi located in the robot. The signal acquisition from the controller was made possible thanks to the implementation of Pydualsense, a package that uses the HID library to capture the controller signals, as well as offering the possibility to control various features of the controller such as LED activation, vibration level, etc.

B. Graphic Interface

The graphical interface refers to the video experience perceived by the user operating the robot. The interface shows the measurements collected by the different sensors, as well as relevant data and information about the system overlaying the video captured by the video camera. The interface was developed using Pygame and OpenCV for its versatility and ease of use.

C. Communication

For the reception and transmission of data between the control station and the robot, a local WiFi network was created in the control station, which handles server and client protocols in the data bus, in addition to this, the use of sockets is proposed for sending and receiving the values obtained from the different sensors.

V. Results

A. Sensor Evaluation

For the evaluation of the integrated sensors, it was proposed to make two measurements in two different conditions, the first one a simulation of disaster environment, where the robot will remain for 25 minutes to take the necessary measurements and a second measurement but in a clear environment under normal conditions for a human being. The results obtained are reflected in Fig. 7 for the measurements in disaster environment and in Fig. 8 for the measurements in normal environment.

The results obtained in a simulated disaster environment show a sustained increase of CO_2 and temperature values over time, which could be reflected in a hypothetical case of disaster, where a potential victim's health would be compromised in a very considerable time interval. The analysis of these data allows the generation of a standard regression curve in which an estimation of the maximum time in which the person would reach critical values is obtained, taking into account the permitted levels of CO_2 in a closed environment, see Table I.

The distance sensors were also evaluated to measure their reliability in providing information to the operator about the perceived distances in the environment. Measurements were made with the sensors as shown in the Table II below.

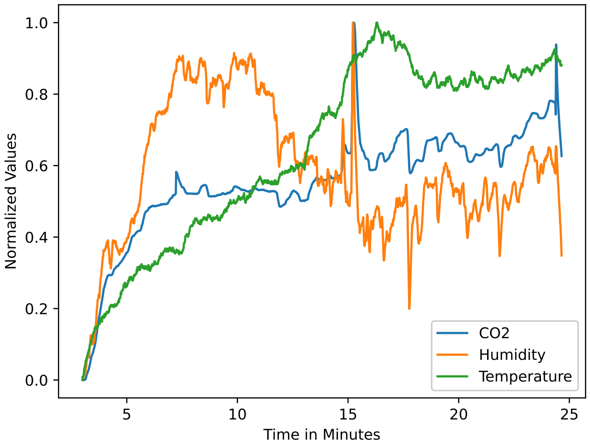


Fig. 7. Measurements taken by TXRob in a Simulated Disaster Environment.

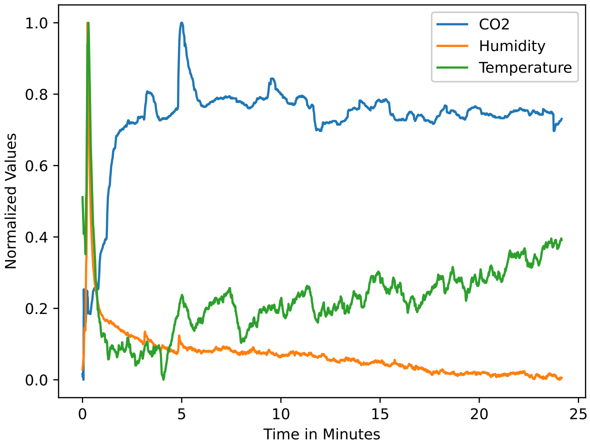


Fig. 8. Measurements taken by TXRob in a Normal Environment.

TABLE II. Measured Distances

Real	100cm	105cm	110cm	115cm	120cm	125cm	130cm
1	101cm	107cm	109cm	114cm	121cm	124cm	130cm
2	101cm	105cm	109cm	113cm	122cm	125cm	131cm
3	100cm	107cm	110cm	112cm	120cm	125cm	130cm
4	101cm	105cm	109cm	113cm	119cm	126cm	128cm
5	99cm	105cm	109cm	114cm	119cm	125cm	130cm
6	100cm	105cm	110cm	112cm	120cm	125cm	131cm
7	99cm	106cm	111cm	113cm	120cm	125cm	129cm
8	100cm	105cm	110cm	114cm	121cm	124cm	129cm
9	98cm	106cm	109cm	115cm	120cm	126cm	132cm
10	100cm	107cm	109cm	115cm	121cm	124cm	131cm

B. Robot Performance

Fig. 9 shows the graph of the time required for the robot to reach the place where the injured person is. From the results, the time required for TXRob to be able to reach the location is directly proportional to the distance it is for flat terrain with no slope. In Fig. 10 the time required for TXRob to reach the site is shown in the same way, but in this test an uneven terrain was considered, with rubble and a slightly steep slope (15 degree slope). It can be seen that for the second test the time increased, but this increase is not so great, so it is not a problem for the robot.

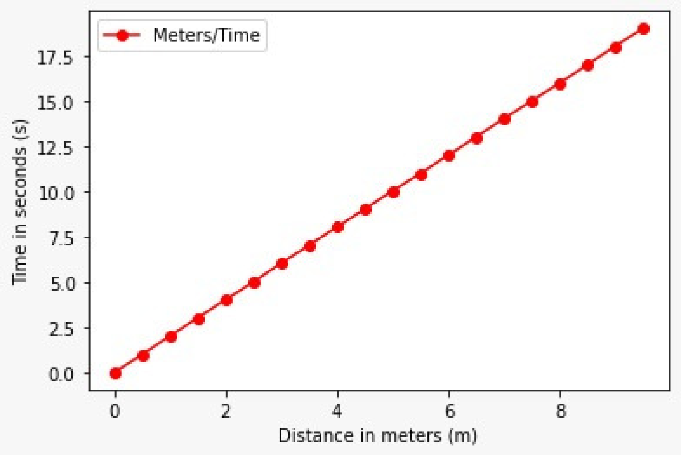


Fig. 9. Plot of the Time required for TXRob to reach the Location on Flat Ground.

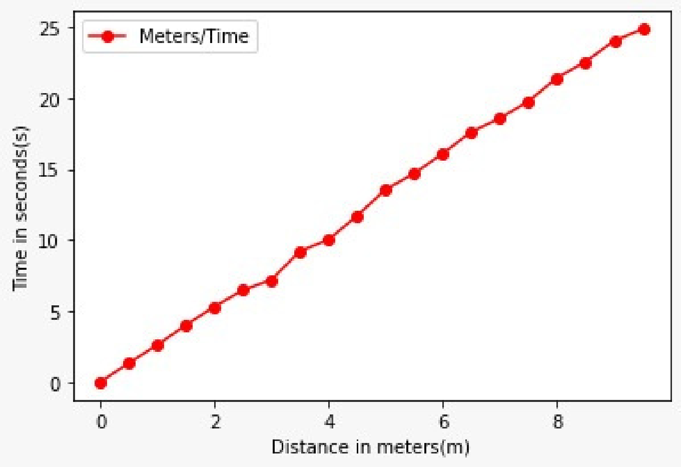


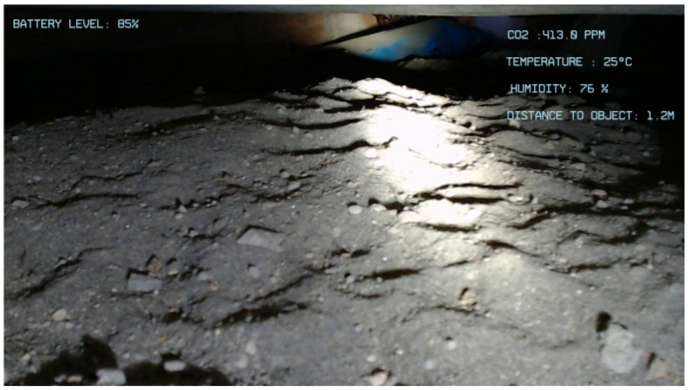
Fig. 10. Plot of the Time required for TXRob to reach the Site on Rough Terrain, with Rubble and a Slight Slope.

C. Display to User

The user display is composed of the video transmitted live by the camera in 720p resolution, together with the measurements obtained in real time by all the built-in sensors. The user display is shown in Fig. 11.

VI. Discussion

Based on the results obtained, the sensors measured and recorded satisfactorily the changes in temperature, humidity and CO_2 when they were subjected to a disaster environment, in the section of Robot Performance, TXRob reached the average speed of 0.47 m/s on a flat surface and 0.36 m/s on a rough surface, making possible the task of exploration in a hypothetical case of disaster where access to possible victims is very reduced due to obstacles, so it is possible to affirm that the measured characteristics of TXRob can fit perfectly in a low-cost robot for exploration and possible rescue work in areas of incidence of natural disasters such as landslides and earthquakes.



BATTERY LEVEL: 85%
CO2 :413.8 PPM
TEMPERATURE : 25°C
HUMIDITY: 76 %
DISTANCE TO OBJECT: 1.2M

Fig. 11. Display Shown to the User.

VII. Conclusion

This paper presents the successful development of a rescue robot for post-disaster scenarios (TXRob), capable of being remotely controlled, i.e. teleoperated. Demonstrating the reliability of the built-in WiFi network based on the Raspberry Pi4 platform. This robot has advantageous features, such as the ability to automatically detect the movement of potentially trapped or endangered people, analyze the environment and determine if it is suitable for rescue agents to enter, and has a suitable design that allows it to move over rough terrain and debris. TXRob also has the ability to avoid strikes by having proximity sensors that alarm the operator, with the results demonstrating the accuracy of the proximity sensors. In addition, its compact structure allows it to access areas inaccessible to humans, which means better exploration of landslides. From the experimental results we can conclude that TXRob can detect different concentrations of gases such as CO_2, NO_2, among others.At the same time that the data obtained by means of the sensors are satisfactorily shown on the display, being the object of future research to be shown in a virtual environment. Finally, it is demonstrated that the robot is able to move through rough terrain in a short time, in summary TXRob is a versatile robot for the rescue of people.

Acknowledgment

The authors would like to thanks Universidad Nacional de San Agustín de Arequipa

References

[1] W. Boerner, ”Future perspectives of SAR polarimetry with applications to multi-parameter fully polarimetric polsar remote sensing & geophysical stress-change monitoring with implementation to agriculture, forestry & aqua-culture plus natural disaster assessment & monitoring within the “pacific ring of fire”,” 2012 IEEE International Geoscience and Remote Sensing Symposium, 2012, pp. 1465-1468, doi: 10.1109/IGARSS.2012.6351258.

[2] J. N. Carpio, F. R. G. Cruz and W. -Y. Chung, ”An earthquake activated power interrupting device using a triaxis accelerometer,” 2016 IEEE Region 10 Conference (TENCON), 2016, pp. 2414-2417, doi: 10.1109/TENCON.2016.7848464.

[3] Instituto Geofisico del Perú. (5 de julio del 2002). ”EL TERREMOTO DE LA REGIÓN SUR DE PERU DEL 23 DE JUNIO DE 2001”. Recuperado de: http://hdl.handle.net/20.500.12816/695

[4] J. A. Heraud and J. A. Lira, ”Study of EQLs in Lima, during the 2007 Pisco, Peru earthquake and possible explanations,” 2011 XXXth URSI General Assembly and Scientific Symposium, 2011, pp. 1-4, doi: 10.1109/URSIGASS.2011.6050739.

[5] Instituto Nacional de Defensa Civil. (26 de mayo del 2019). ”MOVIMIENTO SÍSMICO DE MAGNITUD 8.0 LAGUNAS – LORETO”. Recuperado de: https://onx.la/a72ab

[6] J. A. Heraud, V. A. Centa, P. Mamani, D. Menendez, N. Vilchez and T. Bleier, ”Some Statistical Results from the Triangulation of Electromagnetic Precursors Occurring at the Subduction Zone, Related with Earthquake Activity in Central Peru,” 2021 XXXIVth General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS), 2021, pp. 1-4, doi: 10.23919/URSIGASS51995.2021.9560447.

[7] G. Zhai, W. Zhang, W. Hu and Z. Ji, ”Coal Mine Rescue Robots Based on Binocular Vision: A Review of the State of the Art,” in IEEE Access, vol. 8, pp. 130561-130575, 2020, doi: 10.1109/ACCESS.2020.3009387.

[8] A. Denker and M. C. İşeri, ”Design and implementation of a semi-autonomous mobile search and rescue robot: SALVOR,” 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), 2017, pp. 1-6, doi: 10.1109/IDAP.2017.8090184.

[9] J. P. Queralta et al., ”Collaborative Multi-Robot Search and Rescue: Planning, Coordination, Perception, and Active Vision,” in IEEE Access, vol. 8, pp. 191617-191643, 2020, doi: 10.1109/ACCESS.2020.3030190.

[10] O. SeungSub, H. Jehun, J. Hyunjung, L. Soyeon and S. Jinho, ”A study on the disaster response scenarios using robot technology,” 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), 2017, pp. 520-523, doi: 10.1109/URAI.2017.7992658.

[11] S. Sarkar, A. Patil, A. Hartalkar and A. Wasekar, ”Earthquake rescue robot: A purview to life,” 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2017, pp. 1-7, doi: 10.1109/ICECCT.2017.8118044.

[12] B. Doroodgar, M. Ficocelli, B. Mobedi and G. Nejat, ”The search for survivors: Cooperative human-robot interaction in search and rescue environments using semi-autonomous robots,” 2010 IEEE International Conference on Robotics and Automation, 2010, pp. 2858-2863, doi: 10.1109/ROBOT.2010.5509530.

[13] D. J. Nallathambi, ”Comprehensive evaluation of the performance of rescue robots using victim robots,” 2018 4th International Conference on Control, Automation and Robotics (ICCAR), 2018, pp. 60-64, doi: 10.1109/ICCAR.2018.8384645.

[14] F. Negrello et al., ”Humanoids at Work: The WALK-MAN Robot in a Postearthquake Scenario,” in IEEE Robotics & Automation Magazine, vol. 25, no. 3, pp. 8-22, Sept. 2018, doi: 10.1109/MRA.2017.2788801.

[15] Organización Mundial de la Salud. (22 de septiembre de 2021). Las nuevas Directrices mundiales de la OMS sobre la calidad del aire tienen como objetivo evitar millones de muertes debidas a la contaminación del aire. Recuperado de: https://n9.cl/6fqe7

[16] F. Erden, E. B. Soyer, B. U. Toreyin and A. E. Çetin, ”VOC gas leak detection using Pyro-electric Infrared sensors,” 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 2010, pp. 1682-1685, doi: 10.1109/ICASSP.2010.5495500.

[17] Z. Y. Li, G. Zhang, Z. C. Yang, Y. L. Hao, Y. F. Jin and A. Q. Liu, ”Highly sensitive and integrated VOC sensor based on silicon nanophotonics,” 2017 19th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS), 2017, pp. 1479-1482, doi: 10.1109/TRANSDUCERS.2017.7994338.

[18] D. Cuesta-Frau et al., ”Measuring body temperature time series regularity using approximate entropy and sample entropy,” 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 3461-3464, doi: 10.1109/IEMBS.2009.5334602.

[19] A. Vaz et al., ”Full Passive UHF Tag With a Temperature Sensor Suitable for Human Body Temperature Monitoring,” in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 57, no. 2, pp. 95-99, Feb. 2010, doi: 10.1109/TCSII.2010.2040314.

[20] M. U. H. A. Rasyid, S. Sukaridhoto, A. Sudarsono and A. N. Kaffah, ”Design and Implementation of Hypothermia Symptoms Early Detection With Smart Jacket Based on Wireless Body Area Network,” in IEEE Access, vol. 8, pp. 155260-155274, 2020, doi: 10.1109/ACCESS.2020.3018793.

[21] S. Treratanakulchai and J. Suthakorn, ”Effective vital sign sensing algorithm and system for autonomous survivor detection in rough-terrain autonomous rescue robots,” 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014), 2014, pp. 831-836, doi: 10.1109/ROBIO.2014.7090435.

[22] E. Whitmire, T. Latif and A. Bozkurt, ”Acoustic sensors for biobotic search and rescue,” SENSORS, 2014 IEEE, 2014, pp. 2195-2198, doi: 10.1109/ICSENS.2014.6985475.

[23] S. Jeong, J. Lee, B. Kim, Y. Kim and J. Noh, ”Object Segmentation Ensuring Consistency Across Multi-Viewpoint Images,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 10, pp. 2455-2468, 1 Oct. 2018, doi: 10.1109/TPAMI.2017.2757928.

[24] N. H. Saad, N. A. M. Isa and A. A. M. Salih, ”Local Neighbourhood Image Properties for Exposure Region Determination Method in Nonuniform Illumination Images,” in IEEE Access, vol. 8, pp. 79977-79997, 2020, doi: 10.1109/ACCESS.2020.2990730.

[25] N. Sato, M. Koiji and Y. Morita, ”Quantitative analysis of the relationship between camera image characteristics and operability of rescue robots,” 2016 International Conference on Advanced Mechatronic Systems (ICAMechS), 2016, pp. 533-537, doi: 10.1109/ICAMechS.2016.7813505.

[26] M. A. Andres, L. Pari and S. C. Elvis, ”Design of a User Interface to Estimate Distance of Moving Explosive Devices with Stereo Cameras,” 2021 6th International Conference on Image, Vision and Computing (ICIVC), 2021, pp. 362-366, doi: 10.1109/ICIVC52351.2021.9526934.

[27] Goyzueta, D.V.; Guevara M., J.; Montoya A., A.; Sulla E., E.; Lester S., Y.; L., P.; C., E.S. Analysis of a User Interface Based on Multimodal Interaction to Control a Robotic Arm for EOD Applications. Electronics 2022, 11, 1690. https://doi.org/10.3390/electronics11111690

[28] Hung-Ching Lu and Chih-Ying Chuang, ”The implementation of fuzzy-based path planning for car-like mobile robot,” 2005 International Conference on MEMS, NANO and Smart Systems, 2005, pp. 467-472, doi: 10.1109/ICMENS.2005.119.

[29] Y. Choden, M. Raj, C. Wangchuk, P. Singye and K. Muramatsu, ”Remote Controlled Rescue Robot Using ZigBee Communication,” 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-5, doi: 10.1109/I2CT45611.2019.9033924.

[30] L. Caldas-Calle, J. Jara, M. Huerta and P. Gallegos, ”QoS evaluation of VPN in a Raspberry Pi devices over wireless network,” 2017 International Caribbean Conference on Devices, Circuits and Systems (ICCDCS), 2017, pp. 125-128, doi: 10.1109/ICCDCS.2017.7959718.

Automatic Tariff Classification System using Deep Learning

German Cuaya-Simbro, Irving Hernández-Vera, Elías Ruiz, Karina Gutiérrez-Fragoso
División de Ingeniería en Sistemas Computacionales
Instituto Tecnológico Superior del Oriente del Estado de Hidalgo
Carretera Apan-Tepeapulco Km 3.5 CP 43900, Apan, Hidalgo, México

***Abstract*—The tariff fraction is the universal form of identifying a product. It is very useful because it helps to know the tariff that the product must pay when entering or leaving the country, in this case Mexico. Coffee is a complicated product to identify correctly due to its variants, which at first glance are not distinguishable, which can cause confusion and the tariff to be charged incorrectly. Therefore, the main objective of this project was to develop a system based on Deep Learning models, which allow to identify the tariff code of coffee to import or export this product through the analysis of digital images in real time, generating automatically a general report with this information for the customs broker. The developed system allows speeding up the process of assigning the tariff fraction, and also allows the correct assignment of the tariff fraction, avoiding confusion with other products and the wrong collection of the tariff. It is important to mention that the system, although for the moment it is focused on the country of Mexico, can be used in all customs offices since the tariff fraction is universal. The evaluation of the models was carried out with cross-validation, obtaining an effectiveness of more than 80%, and the tariff fraction assignment model had an effectiveness of 90%.**

***Keywords*—Machine learning; digital image processing; automation process**

I. Introduction

The Harmonized System (HS-code) of Tariff Nomenclature created by the World Customs Organization is widely applied to standardize the exchange of internationally traded goods. The code consists of six digits in general for all countries and in tariff fractions. The tariff fraction is an eight-digit code that represents a good within the Tariff of the General Import and Export Tax (TIGIE). Based on the assignment of the code, the tariff regulations that must be satisfied for the import or export of the goods are established. However, incorrect assignment of the HS-code can result not only in non-compliance with tariff regulations and restrictions, but can also lead to fines, infractions to customs agents or even seizure of the goods [1]. Among the products that are difficult to classify correctly for HS-codes assignment are coffee beans due to the varieties that exist and the different degrees of roasting. In general, coffee beans can be classified into four varieties: Arabica, Excelsa, Liberica and Robusta [2]. However, the classification process for unfamiliar people can be a very complicated task, even for an expert in the domain. This is explained by the fact that the shape and color of the different varieties look similar upon visual inspection. In the field of computational learning, different models have been evaluated to address the automatic classification of coffee beans [3], [4], [5]. This is especially useful for the correct assignment of HS-codes to avoid the negative consequences mentioned above. For this research, a modified basic deep learning architecture [6] was used, in which two layers of feature detection were incorporated. Two databases were integrated, first one consist of 200 images, 50 for each of the main coffee bean varieties: Arabica, Excelsa, Liberica, and Robusta. The second one, include 60 images for three categories of roasted: Not Roasted, Roasted and Dark Roasted. The model built was incorporated into a web-based system to assist in the process of correctly assigning tariff fractions. The performance achieved was 90% accuracy.

The paper is organized as follows: Section II provides a review of studies related to the application of machine learning models for HS-codes assignment as well as works that use this type of models to classify coffee beans with different purposes. Section III describes in detail the methodological strategy applied and the results obtained, as well as their discussion. In the final part, conclusions and future work are presented.

II. Related Work

In the international coffee marketing industry, several studies have been developed to improve the general correct assignment of HS-codes, identification of defects in the coffee beans and also correct class assignment of coffee beans.

In 2015 a paper was published related to the trading system for Singapore Customs, which was based on classifying accurately products and assigning HS-codes based on the text description of the declarations. The technique used was a Background Net. The actual transaction dataset after the pre-processing stage consisted of 40,861 records from chapter 22 and 83,830 records from chapter 90 according to the HS-code system. The data were split 60/40 for the training and testing phases, respectively. The results showed that the model employed had an accuracy greater than 90% for chapter 22 but significantly low for chapter 90 data with accuracy values of around 70% [7]. In another research, several machine learning models were explored to predict HS-code based on commodity descriptions entered by customers. The study followed the cross-industry data mining process methodology. The linear support vector machine model was able to achieve the highest accuracy of 76.3%. The dataset was provided by Dubai Customs through an Artificial Intelligence (AI) hackathon competition held in October 2019. This data consisted of 22,346,194 records where each record had two attributes; the Harmonized System Code (HS-Code) and the description of the user inputs. The machine learning models applied were: Naïve Bayes, K-Nearest Neighbor, Decision

Tree, Random Forest, Linear Support Vector Machine and Adaboost. The authors propose in their study that a hierarchical prediction could be built from the HS-Code header until all the subsections of the HS-Code are identified [8]. The issue of automatic classification of Hormonized System Codes (HS-codes) based on the descriptions provided by the users is also addressed by [9]. Three different Deep Learning architectures were evaluated in the experiment: Hierarchical logistic regression, Neural Machine Translator (NMT) and Long Short Tem Memory (LSTM), the latter two with and without hierarchical loss. Thus, 5 models were analyzed. The dataset consisted of eight months of shipments to a country via the DHL network, which included 1,156 million records. However, records of mask/kn95 shipments and blood samples were excluded so that the results would not be biased by the unusual shipment situation caused by the COVID-19 pandemic. The results showed that the NMT model with hierarchical loss obtained the best performance reaching an accuracy of 85%. In [10] is mentioned that deep learning can be defined as a waterfall that performs non-linear processing to learn multiple levels of data representations. In this context, the study by Lee et al. reports the application of a Deep Learning model to assist in the assignment of HS-codes in collaboration with Korea Customs Service. In the experiment 129, 084 cases were evaluated and the top-3 suggestions made by the model achieve an accuracy of 95.5% [11].

There have also been studies have also been carried out on the topic of coffee bean classification. In [3], measurements in the CIE L*a*b color space are used to characterize green Arabica coffee beans provided by growers in the State of Minas Gerais, Brazil. The algorithms used were a Multi Layer Perceptron feed-forward Artificial Neural Network (ANN) and Naïve-Bayes. The authors report an accuracy of 100%. The data set consisted of 20 samples of 50 grams, 30 images per color were obtained from the following groups: off-white, green, cane green and blue-green. In another study, feed-forward back propagation neural network and K-nearest neighbors algorithms were compared to classify coffee beans from different villages in Cavite, Philippines. The varieties included were Robusta, Excelsa and Liberica. The dataset consisted of 255 images, which were divided into 195 samples for training and 60 for the testing phase. Four morphology characteristics were used in the experiments: area, perimeter, equivalent diameter and percent roundness. The accuracy achieved was 96.66% with ANN and 82.56% with k= 4 for KNN [4]. In [5], 255 samples from the province of Cavite, Philippines were used. The data set included 85 samples per species, considering three species of green coffee: Robusta, Liberica and Excelsa. The images were converted to grayscale to extract morphological characteristics. The experiment included 22 classifier algorithms from 5 families: Decision Trees, Discriminant Analysis, Support Vector Machine, K-Nearest Neighbor and Ensembles. The results showed that Coarse Tree algorithm achieved the best result with 94.1% accuracy. Although in a previous work of the author an accuracy of 96.6% was obtained, it is explained that the Coarse Tree algorithm was faster in the time required for the training phase. Other research aimed to detect and classify Luwak coffee green beans purity into the following purity categories, very low (0-25%), low (25-50%), medium (50-75%), and high (75-100%). The research compared the performance of four pre-trained convolutional neural network (CNN) models: SqueezeNet, GoogLeNet, ResNet-50, and AlexNet. GoogLeNet obtained the best result in training and validation steps and achieve an acuraccy of 89.65% [12]. In the study of Walleign et al., the images of coffee bean samples were collected at Jimma Grading Center in Ethiopia, the beans from the 300 gram sample used for raw evaluation were used to prepare the dataset consisting of 1266 images of coffee beans from 12 quality grades. The dataset was divided into three sets for training (70%), validation (10%) and testing (20%) and accuracy of 89.1% was achieved on the test dataset [13]. In [14], 74 coffee beans of different origins were analyzed and separated into two species: Arabica and Robusta based on their fatty acid composition. The study was based on a Deep Learning approach using a conversion of the raw data into Z-score format. The authors comment that different types of conversion can affect the classification results of coffee species. Statistical analysis and Linear Discriminant Analysis were also used to extract robust features that influenced the Machine Learning process. The author [15] mentions that his article determines the species of coffee bean using the GLCM model (Gray Level Co-Occurrence Matrix) with the help of artificial neural networks. What this author does is create a new method to determine the different species of coffee beans: Arabica, Excelsa and Robusta. 120 images were used for training and 60 images for testing. This author concluded that image processing is effective in determining the quality of coffee bean variants, the ANN classifier had an experimental accuracy of 97.06% which shows that ANN classifier is reliable to classify which species of coffee beans.

On the other hand, several researches have been reported different methods to identification of defects in the coffee beans, for example, he work of Akbar and collaborators was based on the measurement of the quality of green Arabica coffee to classify the beans into five different levels of defects. The extraction of color and texture features of coffee beans was done through color histogram and local binary pattern. The machine learning techniques used were Random Forest and K-Nearest Neighbors, which achieved accuracies of 87.87% and 80.47%, respectively. The experiment included a balanced data set with a total of 900 RGB images for the five classes (defect levels). The data were divided in a 66/33 ratio for the training and testing stages. Authors mentioned that blurred images generated inconsistent values in feature extraction and could lead to incorrect predictions in the classification algorithms [16]. In [17], four models off CNN from Keras Framework were evaluated. The coffee roasting frames were divided into three classes that are Not Yet, Accepted, and Rejected, following the development of coffee bean during the roasting process. The dataset consisted of 4,464 images in Not Yet Class, 3,168 in Accepted, and 3,312 in Rejected. The dataset prepared for deep learning model training had a total of 10,944 images and they were processed into 60% training, 20% validation, and 20% test. The deep learning model training final result showed accuracy >97%. Other research, also focused on the development of more efficient methods to select coffee beans based on shape and particularly, on color descriptors achieved values >88% of accuracy. The techniques applied were Support Vector Machine (SVM), Deep Neural Network (DNN) and Random Forest (RF), to assess coffee beans' defects. Images of each class were taken separately, accounting for a total number of 635 samples (coffee beans) [18]. In other work, an AlexNet-based deep learning model was applied to detect eight types of defects in coffee beans with an accuracy ratio of 95.1%. The original dataset included 3261 samples but by rotation and data augmentation, 7203 images were acquired because the initial set contained few samples with defects [19]. Chou and collaborators proposes a model based on Deep Learning to inspect coffee beans and a Generative-adversarial network (GAN) for Structured Data Augmentation. The proposal aims to contribute to intelligent agriculture with the application of these techniques to remove beans with defects categorized by the SCAA (Specialty Coffee Association of America) and minimize human effort in the process of labeling coffee beans. The model reports an accuracy of over 80% [20]. Previous studies have shown the availability of image processing and machine learning techniques, in [21] is reported the application of techniques such as Convolutional neural networks (CNNs), Support vector machines (SVMs), and k-Nearest-neighbors (KNNs), for the classification of peaberries and normal beans. The separation of peaberries and normal coffee beans increases the value of both peaberries and normal coffee beans in the market. According to the authors, the combination of the CNN and Raspberry Pi 3 holds the promise of inexpensive peaberries and a normal bean sorting system for developing countries. The trained CNN could classify approximately 13.77 coffee bean images per second with 98.19% accuracy of the classification. In [22], an intelligent coffee bean quality inspection system based on deep learning (DL) and computer vision (CV) was developed to assist operators in detecting defects, including mold, fermentation, insect bites, and crushed beans. An open-source dataset of coffee bean images was used for testing. The dataset contains 4626 images of green coffee beans under the same light source, of the total there are 2150 and 2476 images of good and defective beans, respectively. In the experiment, the dataset was divided into a training set of 4000 images and a testing set of 626 images. The accuracy of the ResNet-18 model reached about 93% in the testing set. The student model achieve an accuracy of 79%. The author [23] makes mention in his research that he develops a system which, through convolutional neural networks, is capable of detecting defective coffee beans, for example beans that have some type of imperfection, this author also makes emphasis on that they performed five classifications of defective and non-defective beans better performance with CNN was obtained in all types of defects classification precision was more than 90%.

Finally, we identified some research focused in determine other kind of characteristic of the coffee, flavor and quality, among we can cited the follow: In [24] they mention that the aim of other research is to investigate the feasibility to train machine learning (ML) and deep learning (DL) models for predicting the flavors of specialty coffee using near-infrared spectra of ground coffee as the input. The authors mentioned that effective models provided moderate prediction for seven flavor categories based on 266 samples. The Machine Learning methods applied were Support vector machine and the Deep convolutional neural network (DCNN), which achieved similar performance, with the recall and accuracy being 70–73% and 75–77% respectively . Other study describe a method to classify the geographic origin of coffee beans, comparing popular machine learning methods, including convolutional neural network (CNN), linear discriminant analysis (LDA), and support vector machine (SVM) to obtain the best model. Principal component analysis (PCA) and Genetic algorithm (GA) were applied for LDA and SVM to reduce dimensionality. Ninety-six samples of Arabica coffee beans, representatives of three different geographical origins, were analyzed in the study. The results shown an accuracy of 90% in a prediction set achieved using a CNN method [25]. The author [26] mentions that the evaluation of the color of green coffee beans is an important process to define their quality and price in the market and that this process is normally carried out by means of a visual inspection, this causes some limitations. To solve this, they carry out a system capable of obtaining CIE measurements (Commission Internationale de l'Eclairage) of the coffee beans and in turn classify the beans of this product according to their color. Artificial Neural Networks (ANN) were used as the transformation model and the Bayes classifier was used to classify the coffee beans into four groups: whitish, cane green, green, and bluish-green. The neural networks models achieved a generalization error of 1.15% and the Bayesian classifier was able to classify all samples into their expected classes (100% accuracy).

The review shows that some of the related studies use machine learning models applied to the assignment of HS-codes but not precisely to classify coffee beans. In some cases, the accuracy is not higher than 70%. Experiments using Deep Learning techniques or hierarchical models are also reported, some even considering the top three classification suggestions to increase their performance. However, they either focus on coffee purity categories or defect levels for some variety of coffee beans, or they address only on the roasting process. On the other hand, they have approaches towards intelligent agriculture or the discrimination of peaberries and normal coffee beans, distinction of coffee flavors or identification of the geographical origin of the beans. It is necessary to mention that although one of the works reported 100% accuracy, it is possible that this was due to overfitting because only 20 samples were used. So, we propose the use a CNN model to resolve the problem of this research, classify the variety of coffee beans from digital image analysis. Some researches that report imaging classification from Deep Learning models are: [27] in where the authors make mention that the use of computer vision tools have had a strong impact on the industry given its varied applications and its ability to automate complex and demanding processes. One example of the applications of this kind of tools is the recommendation of clothing in online purchases, up to the characterization and generation of clothing statistics in physical stores. In [28], product recognition can be perceived as a particular research issue related to object detection. However, its application of product image recognition is still less perfect. Thereby, the relevance of our research is to propose a CNN model considering the assignment of tariff fraction for four varieties of coffee beans. There are two important issues to consider in the application of deep learning models. Deep learning algorithms have been used mainly in applications where the data sets were balanced or, as a workaround, in which synthetic data was added to achieve equity. Another concern is that deep learning relies predominantly on large amounts of training data. On the other hand, it is worth noting that the rise of deep learning has been strongly supported by major IT companies (e.g. Google, Facebook, and Baidu) who own a large number of patents in the field and major companies are backed substantially for data collection and processing.

Based on the above, the following differences of this work with those previously described can be highlighted:

- We do not seek to identify imperfections, which focuses on identifying shape features, but the work of [23] gives evidence that it is possible to extract features from images of coffee beans.
- The work of [26], it focuses on the distinction of colors identifiable with the naked eye, which does not happen when images of already toasted coffee beans are analyzed, which is common in a process of importing or exporting coffee in some country, that is, our problem is not reduced to distinguishing different categories of maturation of a coffee bean but to distinguish the variety of coffee from the analysis of a bean with a specific maturation.
- In [15] use an ANN, in our case you propose to take advantage of the advantages provided by the use of a CNN, in addition to including one more kind of coffee.

III. Results

The principal results of this research are reported in four parts, the data set of coffee beans image, the model to identify the variety of coffee beans, the model to determine the tariff fraction and the system to merge the models and generate the documentation related to specific variety of coffee identified.

A. Dataset of Coffee Beans Images

We searched images of coffee beans from several sources of Internet where it specifies the variety showed, and we also obtained real images of coffee beans that we obtained from a coffee beans shop, and then we built two data sets:

Dataset 1, for classifier model of coffee classes: This repository contains images of the four classes of coffee which are: Arabica with 50 images, Excelsa with 50, Liberica with 50 and Robusta with 50. Dataset 2, for classifier model of tariff fraction: The base for this model contains images of three classes of coffee which are: 60 Not Roasted, 60 Roasted and 60 Dark Roasted.

B. Deep Learning Model for Coffee Bean Variety Classification

Deep learning (DL) is a subset of a larger family of data representation-based machine learning algorithms.

The first DL model assigns the variety of coffee that corresponds to a certain image of coffee beans, which can be: Arabica, Excelsa, Liberica or Robusta. For this model, as for the one in the following section, we used the basic deep learning CNN architecture [6] with two feature detection layer. The base model is presented in Fig. 1, which considers input images of 150x150 pixels. The first convolution layer expands the information to 32 levels deep for convolutions, pooling reduces the visual information to a quarter of its size (75×75 pixels). 6x6 size filters were used. In the next stage, the information for convolution was duplicated, although with a smaller filter (4x4). Pooling was applied again, reducing the information from the previous layer to a quarter of its size (37×37 pixels). After that, the flattening layer was applied and the information was reduced to a vector of size 256. Finally, a Softmax layer was applied to describe the probabilities for each of the four classes (variety of coffee) to be predicted. Some changes were made due to the fact that satisfactory results were not obtained with the base architecture. Two layers were modified, one of convolution and one of reduction, in the same way an Adam optimizer was chosen, which is within the classification layer, this optimizer was chosen since it has a more effective error reduction. More details of the layers of the architecture used for the developed deep learning model are illustrated in Table I.

TABLE I. Summary of the Architecture used to Built Deep Learning Model for Coffee bean Variety Classification

Specs
Convolution layer (32) + Filter (6,6) + Height, Length (150)
Pooling
Convolution layer (64) + Filter (4,4)
Pooling
Classification layer (ReLu), 256 neurons, Softmax, Adam optimized

In more general terms, the first convolution layer will process images from coffe beans with a certain height and length (it will be resized to 150x150 px), and this layer will allow us to detect basic characteristics such as curves or lines, basic textures and so on from the images. For the first reduction layer we say that it will have a maxpooling and a certain pool size, this layer reduces the first convolution layer in order to gain some scale invariance in the input images. For the second convolution layer it is basically the same as the first one, except that in this image we will no longer assign a height, length, nor will it have any activation function. For the second layer of reduction we say that it will have a maxpooling and a 37×37 size of pool, to again, gain some extra invariance. This layer reduces the second layer of convolution. The classification layer contains a property which makes the image, which is now very deep and very small, will now become one-dimensional that contains all the information of our neural network, it also contains 256 neurons and the ReLu activation function, it does a dropout process which makes 50% of these neurons activate at each step (number of times the information is processed), this is done so that it not only learns a specific way to classify coffee and can adapt to new information (some kind of overfitting prevention) and the number of coffee classes which the model will contain in this case are four. Finally, this model in order to classify the variety of coffee beans required 22467300 parameters. In the two convolutions the number of parameters was 3488 and 32832 respectively. Once the network is flattened, the number of parameters to learn is 22429952. In the last classification layer, only 1028 parameters are required in order to detect four classes. A decimation function was performed to fit the images to the model input of $150 \times 150 \times 3$ pixels.

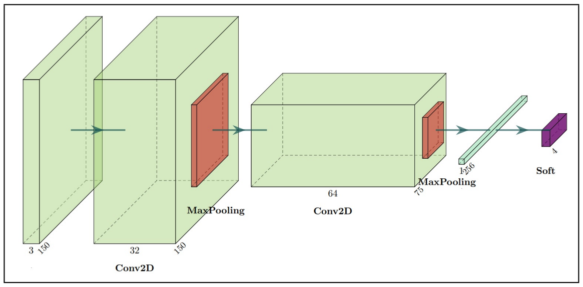


Fig. 1. Base Architecture of the Model used to Estimate the Four Classes of Coffee Beans. The Architecture that Estimates the Roasting Levels is Similar, having Three Classes in the Last Layer Instead of Four. More Details in the Tables I and II.

C. Deep Learning Model for the Identification of Tariff Fraction

The goal of this model is to classify the coffee into the following classes: Not Roasted, Roasted and Dark Roasted, once the coffee is classified within any of these, the model will assign a tariff fraction which is obtained from a database and it will generate detailed information of the product. In contrast as the previous model, the base architecture of CNN model was used, like show the Table II (note that, it is very similar to the corresponding architecture presented in Fig. 1). This is because, in this case, the classification focuses on color identification. In this second model, only the first convolution layer changes, reducing its number of parameters from the first convolution to 1568 (the first contains 3488). The final number of parameters therefore in this model was 22465380 parameters. These 22 million parameters were trained using a traditional Adam optimizer based on the Tensorflow library in the Python programming language. Since the training images consisted of a reduced dataset (due to the few examples found in the literature, even including our own examples), the model was not expensive to train locally. The effectiveness of the model (Section III-D) shows the results using a cross-validation technique, which could be done efficiently due to the few image examples mentioned above.

TABLE II. Summary of the Architecture used to Built Deep Learning Model for the Identification of Tariff Fraction

Specs
Convolution layer (32) + Filter (4,4) + Height, Length (150)
Pooling
Convolution layer (64) + Filter (4,4)
Pooling
Classification layer (ReLu), 256 neurons

D. Models Effectiveness Report

To validate the performance of the models we used 10 cross-validation, because the size of datasets, thereby we used 80% of data set to train the model and 20% for validation; this was applied to both models.

The obtained results (averages) are showed in Tables III and IV, performance of the model to identify the coffee class and performance of the model to identify the tariff fraction, respectively. The tables present precision, recall and f1-score measures, as well as the micro, macro and weighted measures.

TABLE III. Performance of the Coffee Variety Classifier Model

	Averages		
Class	Precision	Recall	f1-score
Arabica	72.70%	80.00%	75.60%
Excelsa	79.90%	90.00%	83.80%
Liberica	91.00%	78.00%	83.90%
Robusta	100.00%	90.00%	94.50%
		Macro	Weighted
Precision		86.00%	86.00%
Recall		84.60%	84.10%
F1-Score		84.30%	84.40%
Accuracy (Micro avg)	84.10%		

TABLE IV. Performance of the Tariff Fraction Classifier Model

	Averages		
Class	Precision	Recall	f1-score
Not Roasted	98.30%	98.00%	98.00%
Roasted	87.20%	88.00%	86.30%
Dark Roasted	89.20%	84.00%	85.40%
		Macro	Weighted
Precision		91.40%	91.40%
Recall		90.00%	90.00%
F1-Score		89.80%	89.80%
Accuracy (Micro avg)	90%		

The results of the Table III show that the model has a greater error when identifying the Arabica class, this is probably due to the fact that it has too much similarity, visually, to the Excelsa class. Fig. 2 shows the confusion matrix obtained for the results in Table III. The matrix shows more clearly the errors between the Arabica and Excelsa classes.

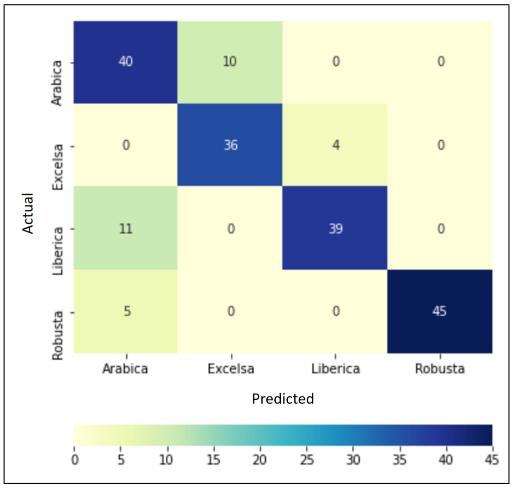
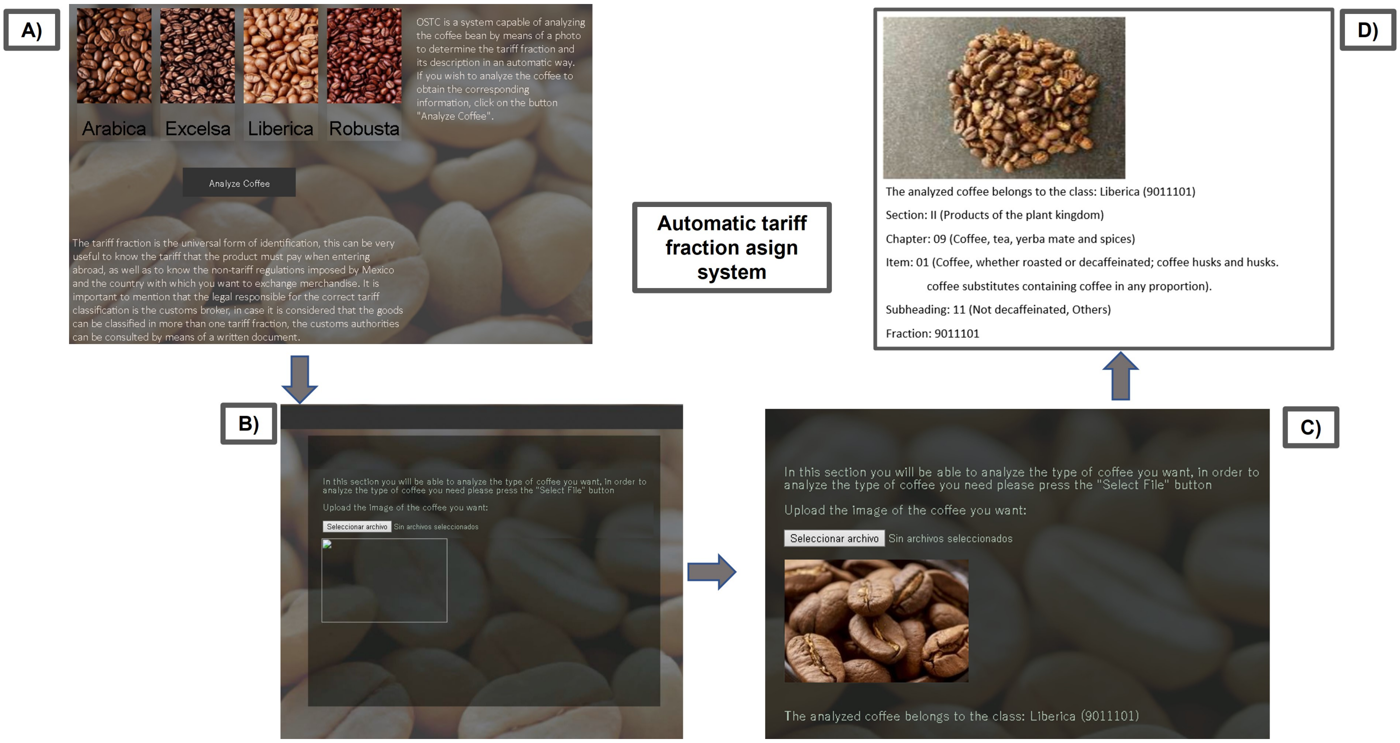


Fig. 2. Confusion Matrix of the Four Classes of Coffee Beans recognized by the Model. The Results show that Arabica Coffee is more difficult to Distinguish and is Confused with Liberica and Excelsa. Robusta is the Easiest Class of Coffee to Distinguish by the Model.

On the other hand, due to we have balanced classes, the measures micro, macro and weighted were similar, and we can take the micro measure as the Global Accuracy of the model. Thereby, the model have a global accuracy greater than 80% and the performance measured of each class show that the model is capable to distinguish the coffee beans class with good performance despite of visually them have the same color and shape.


A)
Arabica
Excelsa
Liberica
Robusta
Analyze Coffee
B)
Seleccionar archivo
Sin archivos seleccionados
Automatic tariff fraction asign system
C)
Seleccionar archivo
Sin archivos seleccionados
D)

OSTC is a system capable of analyzing the coffee bean by means of a photo to determine the tariff fraction and its description in an automatic way. If you wish to analyze the coffee to obtain the corresponding information, click on the button "Analyze Coffee".

The tariff fraction is the universal form of identification, this can be very useful to know the tariff that the product must pay when entering abroad, as well as to know the non-tariff regulations imposed by Mexico and the country with which you want to exchange merchandise. It is important to mention that the legal responsible for the correct tariff classification is the customs broker, in case it is considered that the goods can be classified in more than one tariff fraction, the customs authorities can be consulted by means of a written document.

In this section you will be able to analyze the type of coffee you want, in order to analyze the type of coffee you need please press the "Select File" button

Upload the image of the coffee you want:

In this section you will be able to analyze the type of coffee you want, in order to analyze the type of coffee you need please press the "Select File" button

Upload the image of the coffee you want:

The analyzed coffee belongs to the class: Liberica (9011101)

The analyzed coffee belongs to the class: Liberica (9011101)

Section: II (Products of the plant kingdom)

Chapter: 09 (Coffee, tea, yerba mate and spices)

Item: 01 (Coffee, whether roasted or decaffeinated; coffee husks and husks. coffee substitutes containing coffee in any proportion).

Subheading: 11 (Not decaffeinated, Others)

Fraction: 9011101

Fig. 3. Sequence and Interfaces of Web System Usage. A) Main Interface, where a request of an Image to Analyse is Performed, B) Interface to Upload the Image to analyse by the Deep Learning Models, C) Interface where the System Presents the Class of Coffee Beans and the Tariff Fraction associated to them, and D) Final Report generated automatically by the System.

The results of the Table IV show that the model has a greater error when identifying the Roasted class, according to the performance measures for class, this is probably due to the fact that the roast sometimes has an identical tone to the dark roast and therefore the model becomes confused. Fig. 4 shows the confusion matrix for the results of Table IV. The difficulty in recognizing coffee beans with “roasted” and “dark roasted” levels can be seen more clearly.

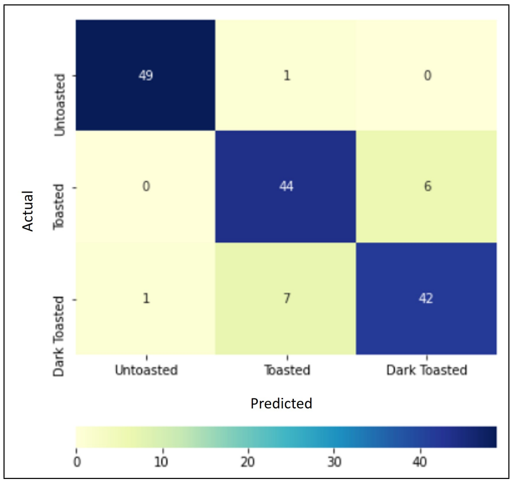


Fig. 4. Confusion Matrix of the Three Levels of Roasting. As Might be Expected, Roasted and Dark Roasted Classes are more difficult to Distinguish Since they have Similarities in Visual Appearance if Particular Coffee Beans are taken. These Results Reflect Visual Consistency for Coffee Roasting Levels.

In the same way that the other of our models, we can consider the micro measure like the global accuracy, 90%. This result is comparable to the similar work of [26] where their objective is to identify different types of coffee beans according the color of them, and they reported a global performance of 97.06%.

E. Web System

Finally, a web system was built, where the created Deep learning models were integrated, this system automatically assigns the tariff fraction of coffee beans from the analysis of a digital image, that image is entered into the first model and this determines the variety of coffee, then the second model determines if the coffee beans are roasted, dark roasted or not roasted in order to assign the corresponding tariff fraction. Finally, a document is automatically generated which contains the image uploaded by the user and the tariff item broken down in detail. The Fig. 3 shows images of the interfaces system and the general process to use it.

F. Discussion

This article presented the structure and model of a system that allows determining types of coffee beans as well as the roasting level in order to assign the tariff fraction as support in customs and, therefore, usable by customs brokers. It was noted that there is little information on visual datasets about coffee beans. Mostly concrete examples exist but not *many* image examples. This implied an interest in generating a small compendium of images generated by the authors of this work that allowed strengthening the results of the computer vision models presented. Computer vision models are based on Deep Learning techniques that, unlike the state of the art, present a lighter architecture in order to achieve finer recognition of coffee beans patterns that are usually difficult to identify with the naked eye. With the accuracy obtained from the models, the results were used to estimate the tariff fraction. These results, taken to a web system, facilitate the manual task of estimating said tariff fraction by people who are not familiar with agronomy and, in particular, with coffee beans varieties, or their roasting level. In addition, even when the main goal for this project is to facilitate the process of entering and leaving coffee from our country, since the system is web-based, it can be used ubiquitously by a diverse number of users, requiring only an Internet connection.

IV. Conclusion and Future Work

We have were present a novel system to automatically distinguish the variety of coffee from the analysis of a bean with a specific maturation, Arabica, Excelsa, Liberica, and Robusta, with a accuracy gather than of 80%, that is relevant in a process of importing or exporting coffee in some country, which helps to avoid the negative consequences of incorrect classification. Which is one of the main differences with similar works, given that these works only report the development of machine learning models but do not report their implementation in any tool that is easy to use by end users.

So, we described the deep learning models built and the accuracy of them, 84.1% for model for coffee bean variety classification and 90% by model the identification of tariff fraction respectively. These results are comparable with research in the state of the art. In this way, this results support the feasibility of the application of these model to the tariff fraction assign process.

The base architecture used for the development of Deep Learning models was very useful since not many changes had to be made for model classify of coffee, and no changes were done for the tariff fraction assign model.

The assessment made by end users was good and they consider that the system meets the needs of customs agents. In addition to the fact that the system is fast and innovative, so we consider that we not only obtained Deep Learning models, but we integrated them to a web system for use them in a simple way.

Finally, and as a future work, we consider to improve the performance of both model, but specifically the classify of coffee model, first increasing the size of the datasets, and second, carrying out other changes in the CNN architecture base used, for example, add a third convolutional layer, or even to probe other architectures like LeNet or Alexnet.

References

[1] C. V. Espinoza Zhingre, "Incidencia de la falta de una correcta clasificación arancelaria para la comercialización internacional de las mercancías," Master's thesis, Machala: Universidad Técnica de Machala, 2015.

[2] I. Ismail, M. S. Anuar, and R. Shamsudin, "Physical properties of liberica coffee (coffea liberica) berries and beans," *Pertanika Journal of Science and Technology*, vol. 22, pp. 65–79, 01 2014.

[3] E. M. de Oliveira, D. S. Leme, B. H. G. Barbosa, M. P. Rodarte, and R. G. F. A. Pereira, "A computer vision system for coffee beans classification based on computational intelligence techniques," *Journal of Food Engineering*, vol. 171, pp. 22–27, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0260877415300108

[4] E. Arboleda, A. Fajardo, and R. Medina, "Classification of coffee bean species using image processing, artificial neural network and k nearest neighbors," in *2018 IEEE International Conference on Innovative Research and Development*, ser. ICIRD 2018, 2018, pp. 1–5.

[5] E. Arboleda, "Comparing performances of data mining algorithms for classification of green coffee beans," *International Journal of Advanced Technology and Engineering Exploration*, vol. 8, pp. 1563–1567, 2019.

[6] M. Matlab, "Introducing deep learning with matlab," Mar 2017. [Online]. Available: https://www.mathworks.com/campaigns/offers/deep-learning-with-matlab.html

[7] L. Ding, Z. Fan, and D. Chen, "Auto-categorization of hs code using background net approach," *Procedia Computer Science*, vol. 60, pp. 1462–1471, 2015, knowledge-Based and Intelligent Information & Engineering Systems 19th Annual Conference, KES-2015, Singapore, September 2015 Proceedings. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915023510

[8] F. Altaheri and K. Shaalan, *Exploring Machine Learning Models to Predict Harmonized System Code*. Springer Nature Switzerland, 04 2020, pp. 291–303.

[9] X. Chen, S. Bromuri, and M. van Eekelen, "Neural machine translation for harmonized system codes prediction," in *2021 6th International Conference on Machine Learning Technologies*, ser. ICMLT 2021. New York, NY, USA: Association for Computing Machinery, 2021, p. 158–163. [Online]. Available: https://doi.org/10.1145/3468891.3468915

[10] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–36, 2018.

[11] E. Lee, S. Kim, S. Kim, S. Park, M. Cha, S. Jung, S. Yang, Y. Choi, S. Ji, M. Song, and H. Kim, "Classification of goods using text descriptions with sentences retrieval," 2021. [Online]. Available: https://arxiv.org/abs/2111.01663

[12] Y. Hendrawan, S. Widyaningtyas, and S. Sucipto, "Computer vision for purity, phenol, and ph detection of luwak coffee green bean," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 17, p. 3073, 12 2019.

[13] S. Wallelign, M. Polceanu, T. Jemal, and C. Buche, "Coffee grading with convolutional neural networks using small datasets with high variance," *Journal of WSCG*, vol. 27, 01 2019.

[14] Y.-C. Hung, F.-S. Lee, and C.-I. Lin, "Classification of coffee bean categories based upon analysis of fatty acid ingredients," *Journal of Food Processing and Preservation*, vol. 45, no. 9, p. e15703, 2021.

[15] A. M. Castillo, R. D. Aradanas, E. R. Arboleda, A. A. Dizon, and R. M. Dellosa, "Coffee type classification using gray level co-occurrence matrix feature extraction and the artificial neural network classifier," *International Journal of Scientific & Technology Research*, vol. 8, pp. 2277–8616, 2019.

[16] M. N.S. Akbar, E. Rachmawati, and F. Sthevanie, "Visual feature and machine learning approach for arabica green coffee beans grade determination," in *2020 the 6th International Conference on Communication and Information Processing*, ser. ICCIP 2020, 2020, p. 97–104.

[17] M. Hakim, T. Djatna, and I. Yuliasih, "Deep learning for roasting coffee bean quality assessment using computer vision in mobile environment," in *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2020, pp. 363–370.

[18] F. Santos, J. Rosas, R. Martins, G. Araújo, L. Viana, and J. Gonçalves, "Quality assessment of coffee beans through computer vision and machine learning algorithms," *Coffee Science - ISSN 1984-3909*, vol. 15, p. e151752, Aug 2020. [Online]. Available: http://www.coffeescience.ufla.br/index.php/Coffeescience/article/view/1752

[19] S.-J. Chang and C.-Y. Huang, "Deep learning model for the inspection of coffee bean defects," *Applied Sciences*, vol. 11, 2021.

[20] Y.-C. Chou, C.-J. Kuo, T.-T. Chen, G.-J. Horng, M.-Y. Pai, M.-E. Wu, Y.-C. Lin, M.-H. Hung, W.-T. Su, Y.-C. Chen, D.-C. Wang, and C.-C. Chen, "Deep-learning-based defective bean inspection

with gan-structured automated labeled data augmentation in coffee industry," *Applied Sciences*, vol. 9, no. 19, 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/19/4166

[21] H. Gope and H. Fukai, "Peaberry and normal coffee bean classification using cnn, svm, and knn: Their implementation in and the limitations of raspberry pi 3," *AIMS Agriculture and Food*, vol. 7, pp. 149–167, 03 2022.

[22] P. Wang, H.-W. Tseng, T.-C. Chen, and C.-H. Hsia, "Deep convolutional neural network for coffee bean inspection," *Sensors and Materials*, vol. 33, no. 7, pp. 2299–2310, 2021.

[23] H. Fukai, J. Furukawa, C. Pinto, and C. Afonso, "Classification of green coffee beans by convolutional neural network and its implementation on raspberry pi and camera module," *Timorese Academic Journal of Science and Technology*, 2018.

[24] Y.-T. Chang, M.-C. Hsueh, S.-P. Hung, J.-M. Lu, J.-H. Peng, and S.-F. Chen, "Prediction of specialty coffee flavors based on near-infrared spectra using machine- and deep-learning methods," *Journal of the Science of Food and Agriculture*, vol. 101, no. 11, pp. 4705–4714, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jsfa.11116

[25] S. Yang, C. Li, Y. Mei, W. Liu, R. Liu, W. Chen, D. Han, and K. Xu, "Determination of the geographical origin of coffee beans using terahertz spectroscopy combined with machine learning methods," *Frontiers in Nutrition*, vol. 8, 2021. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnut.2021.680627

[26] E. M. de Oliveira, D. S. Leme, B. H. G. Barbosa, M. P. Rodarte, and R. G. F. A. Pereira, "A computer vision system for coffee beans classification based on computational intelligence techniques," *Journal of Food engineering*, vol. 171, pp. 22–27, 2016.

[27] S. R. Sepúlveda Osses, "Detección de prendas de vestir utilizando modelos de detección de objetos basados en deep learning," Master's thesis, Universidad de Chile, 2020.

[28] X.-S. Wei, Q. Cui, L. Yang, P. Wang, and L. Liu, "Rpc: A large-scale retail product checkout dataset," 2019. [Online]. Available: https://arxiv.org/abs/1901.07249

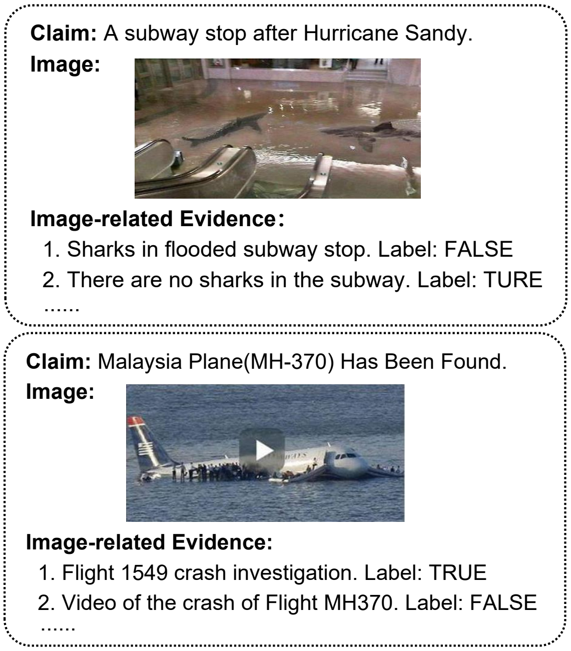
Fake News Detection in Social Media based on Multi-Modal Multi-Task Learning

Xinyu Cui
Northeast Forestry University
Harbin, China

Yang Li*
Northeast Forestry University
Harbin, China

***Abstract*—The popularity of social media has led to a substantial increase of data. The task of fake news detection is very important, because the authenticity of posts cannot be guaranteed. In recent years, fake news detection combining multi-modal information such as images and videos has attracted wide attention from scholars. However, the majority of research work only focuses on the fusion of multi-modal information, while neglecting the role of external evidences. To address this challenge, this paper proposes a fake news detection method based on multi-modal and multi-task learning. When learning the representation of the news posts, this paper models the interaction between images and texts in posts and external evidences through a multi-level attention mechanism, and uses evidence veracity classification as an auxiliary task, so as to improve the task of fake news detection. Authors conduct comprehensive experiments on a public dataset, and demonstrate that the proposed method outperforms several state-of-the-art baselines. The ablation experiment proves the effectiveness of the auxiliary task of evidence veracity in fake news detection.**

***Keywords*—Multi-modal fake news; multi-task learning; external evidences; multi-level attention mechanism**



Claim: A subway stop after Hurricane Sandy.
Image:

Image-related Evidence:
1. Sharks in flooded subway stop. Label: FALSE
2. There are no sharks in the subway. Label: TURE
......

Claim: Malaysia Plane(MH-370) Has Been Found.
Image:

Image-related Evidence:
1. Flight 1549 crash investigation. Label: TRUE
2. Video of the crash of Flight MH370. Label: FALSE
......

Fig. 1. Two Examples of Multi-Modal Fake News. The Claim is the Text Information of Multi-Modal News, the Image is the Visual Information Contained in the Multi-Modal News, and the Evidence is the Web Pages Extracted from Google.

I. Introduction

Social media is an important platform for people to share and obtain information, and has become an indispensable part of people's daily life. But at the same time, the characteristics of easy access and manipulation of social media information also promote the proliferation of fake news. Fake news on social media not only affects public opinion, but also does serious harm to the economy [1], politics [2], public health [3] and society. Therefore, fake news detection has become an important research issue.

The purpose of fake news detection is to automatically determine whether the statements in news posts are true or false. Some news posts contain videos or images besides words, which are more attractive and deceptive than textual news [4]. According to statistics, the average forwarding times of posts containing images are about 11 times that of posts without images [5]. Multi-modal fake news usually contains some distorted or confusing images [6]. As shown in Fig. 1, the upper image is obviously processed by tools, while the image in the lower is a misleading image that is inconsistent with the text.

Recent works have made lots of attempts on multi-modal fake news detection [7]. Some researches simply combine textual features with visual features to obtain multi-modal features [8]. Wang et al. [9] use Text-CNN and VGG-19 to extract text and image features respectively, and then simply concatenate them to classify the news. Sing et al. [10] manually design textual and visual features from four dimensions: content, organization, emotion and manipulation, and then concatenate them to detect fake news. In order to capture the interactions of multi-modal features, Wu et al. [11] stack multiple co-attention layers to fuse the multi-modal features. Qi et al. [12] extract three kinds of text-image correlations to capture multi-modal clues. However, the above methods only use the information of the news itself and neglect the use of external evidence.

To this end, this paper proposes a fake news detection method via Multi-modal and Multi-task Learning (MML). Different from previous studies, the classification of evidence veracity is used as an auxiliary task of fake news detection. MML first extracts the features of the image by a multi-layer CNN model, and then obtains evidence representations through claim-evidence correlation representation learning. Finally, the representations of image and image-related evidence are fused through the co-attention mechanism. Specifically, this paper jointly trains fake news detection and evidence classification, the two tasks share the representation of evidence.

The main contributions of the paper are as follows:

1) This paper proposes an end-to-end neural network to detect multi-modal fake news on social media by simultaneously learning the deep correlations between the image, claim and the evidence.

2) This paper extracts the image-related evidence and improves the performance of fake news detection through a multi-task learning framework.

3) Authors design detailed experiments to prove the effectiveness of the proposed model, and verify the effectiveness of multi-modal learning and multi-task learning in this task.

The remaining sections of the paper are structured as follows: Section II introduces the literature survey in the field of fake news detection and multi-task learning. After that, Section III explains the methodology. Then, Section IV describes the results and discussion followed by Section V conclusion and future enhancements.

II. Literature Survey

This section briefly summarizes the existing work in the field of fake news detection and multi-task learning.

A. Fake News Detection

For news that only contains texts, besides the text information, the propagation structure of news on social networks is commonly used to detect fake news. Liu et al. [13] presented a kernel graph attention network, which performed more fine-grained fact verification based on kernel-based attentions. Zhong et al. [14] applied semantic role labeling to parse each evidence sentence and established links between arguments to build a graph structure for information detection. Different from the graph structure constructed in the above methods, Ma et al. [15] and Bian et al. [16] modelled the propagation of posts on the Weibo platform by tree structures. Some researchers have different opinions about the research direction of fake news. They think it is very important to study the interpretability of fake news detection. Shu et al. [17] developed a joint attention graph to capture the top K interpretable sentences and user comments. Wu et al. [18] proposed a dual-view model based on collective cognition and individual cognition for interpretative claim verification.

While for multi-modal news, various methods have been proposed to utilize the multi-modal information and detect fake news. Vo et al. [19] proposed to use images as a supplement to news content, and used text matching layer and visual matching layer to detect text and images respectively. Jin et al. [20] treated each image or video as a topic and used the credibility of these topics as a new feature to detect fake news. However, a key problem with using multi-modal information for fake news detection is that the multi-modal information usually comes from another real event, and the content seems to correspond to the text in the fake news. At this point, although the image itself is real, it does not actually match the text content. The above methods ignore this problem and do not fully integrate text and multimedia content. Based on this, Wu et al. [11] proposed a joint multi-modal attention network to integrate the text features and visual features of fake news. Qi et al. [12] captured the correlation between text features and visual features by extracting three kinds of text-image features. However, these methods ignore the use of external evidence. Wen et al. [21] leveraged the semantic similarity between news and external evidence to capture the mismatch between text content and multi-modal information. However, this method did not fuse the physical features of image. To overcome the above limitations, this paper proposes a method to capture the physical features of image, and learns the deep correlations between the image, claim and the evidence.

B. Multi-Task Learning

Multi-task learning refers to the joint learning of related tasks that share representation information, so that these tasks can achieve better results than training a single task. In recent years, multi-task learning has been proved to be effective in various NLP tasks, including fake news detection. Kochkina et al. [22] constructed a multi-task learning framework consisting of three tasks: veracity classification, stance classification and rumor detection. The proposed method was represented by a shared LSTM layer (hard parameter sharing), followed by many task-specific layers. Ma et al. [23] jointly modelled rumor detection and stance classification by using two RNN-based architectures with shared layers. Wu et al. [24] explored a sharing layer of gate mechanism and attention mechanism, which can selectively capture valuable sharing features for fake news detection and stance detection. Li at al. [25] proposed a neural network model for multi-task learning of rumor detection and stance classification, including a shared layer and two task-specific layers. However, all the above multi-task learning methods are based on the joint training of fake news detection and stance detection. To the best of author's knowledge, this paper makes the first attempt to jointly model fake news detection and evidence veracity classification in multi-task learning.

III. Methodology

A. Overview

This paper proposes a Multi-modal and Multi-task Learning method for fake news detection (MML). As shown in Fig. 2, the proposed model mainly contains three parts: visual representation learning, textual representation learning, and fake news classification.

The problem definition is as follows. Suppose that $P = (p_1, \ldots, p_n)$ is a set of multi-modal news posts from social media, the text in the news is denoted as a claim C_j where $j \in [1, n]$. $E_i = \{e_i^1, e_i^2, \ldots, e_i^m\}$ is a set of evidence for news p_i, composed of the titles of web pages searched from Google. Given a news post p_i and the corresponding evidence set E_i, the main task aims to predict whether p_i is a fake news based on its multi-modal representation learned from MML. For the task of evidence veracity classification, in the training stage, this paper uses the label of evidence to learn the representation of evidence and shares it with the main task.

B. Visual Representation Learning

Since the fake-news images are often re-compressed images or tampered images, they are different from real-news

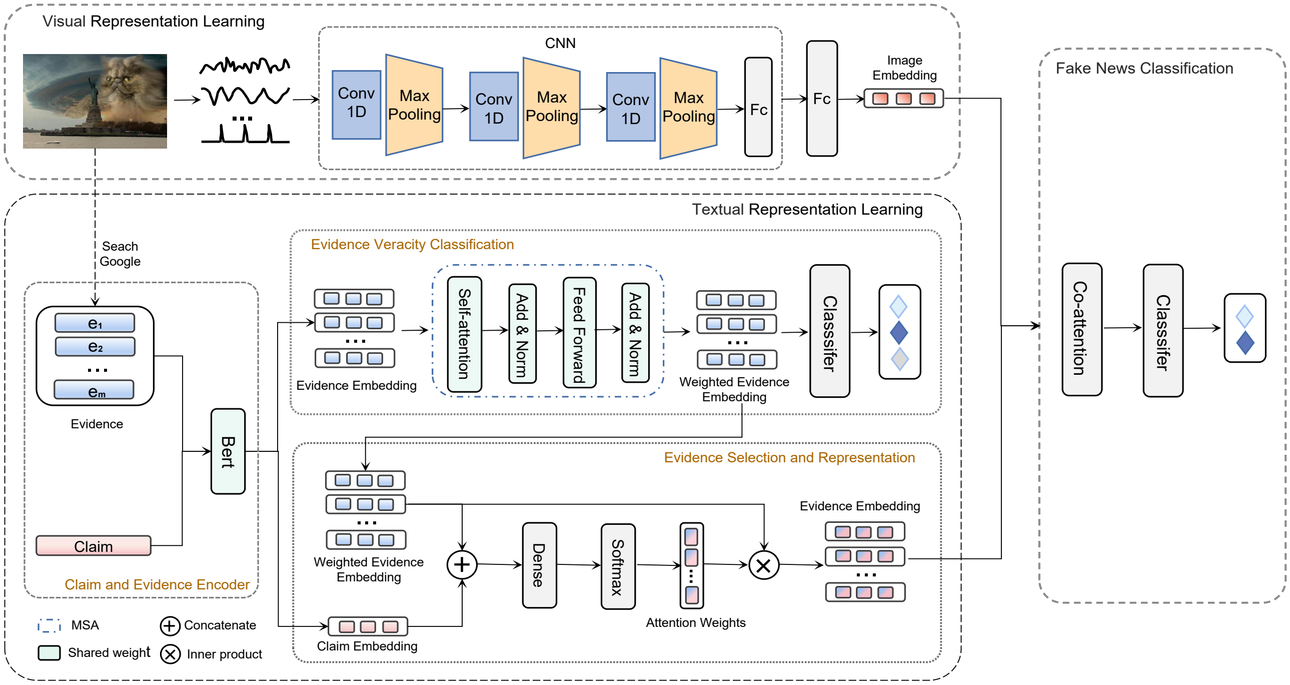


Fig. 2 Illustration of the Proposed Model MML.

images in frequency domain, which are usually periodic. Inspired by Qi et al. [26], the discrete cosine transform (DCT) is first used to transform the image in the news post from spatial domain to frequency domain, to obtain 64 hisograms, which can be represented by 64 vectors $V_0, V_1, \ldots, V_{63}$ with a fixed size. After that, this paper feeds each vector to the multi-layer CNN model consisting of three convolution blocks and a fully connected layer, where each convolution block contains a one-dimensional convolution layer and a max-pooling layer. Finally, a fully connected layer with ReLU activation function (denoted as “Fc” in Fig. 2) is added to get the feature representation of image R_v.

C. Textual Representation Learning

To capture the correlations of the semantics and visual information of the news posts, MML extracts image related web pages from Google to serve as the evidence of the claim. At the same time, in order to make a selection of evidence, MML uses the evidence veracity classification task to assist the fake news detection task. In this part, the claim and the evidence are first fed into a BERT-based encoder, then through the evidence veracity classification task, the importance of evidence is learned. Finally, the textual representation is learned based on the co-attention of claim and the evidence.

1) Claim and Evidence Encoder: This paper uses BERT to obtain the representations of claim and its corresponding m related evidences. The BERT model is a bidirectional coding representation model based on the transformer structure proposed by Devlin et al. [27]. Compared with the traditional Recurrent Neural Network (RNN) and Long Short-Term Memory networks (LSTM) used for NLP tasks, the transformer structure is more powerful in encoding texts. It consists of six encoder-decoders stacked with the same structure. Each encoder consists of two sub-layers, i.e., a feedforward layer and a multi-head attention layer, and each decoder consists of three sub-layers: a feedforward layer, a multi-head attention layer and a masked multi-head attention layer. In addition, add and normalization functions are added to each sub-layer. The BERT model achieves better performance in existing models by stacking twelve-layer Transformer Encoders.

Given a claim C and a set of evidences $E = \{e_1, e_2, \ldots, e_m\}$ corresponding to the claim, BERT model is used to generate the representations of the claim and each evidence:

$$R_c = BERT(C) \tag{1}$$

$$h_j = BERT(e_j) \tag{2}$$

where e_j is the *j*-th evidence corresponding to the claim C. Next, the total evidence representation H is obtained by concatenating the representation of each evidence:

$$H = h_1 \oplus h_2 \oplus \cdots \oplus h_m \tag{3}$$

where m represents the number of evidence corresponding to the claim C.

2) Evidence Veracity Classification: Since there are a lot of web pages searched from Google, it is of great significance for fake news detection that how to find the “useful evidences” and make use of them. Taking the evidence representation H as input, the Transformer encoder is used to capture the correlations of evidences, and a Multi-layer Perceptron (MLP) is used to classify the evidence into three pre-defined categories: True, False and Unverified.

The objective of evidence veracity classification task is to minimize the cross-entropy loss function:

$$\mathcal{L}_e = -\sum_{i=1}^{m} q_i log p_i \tag{4}$$

where p_i denotes predicted probability of evidence i, q_i refers to the ground-truth label of evidence i. By classifying the veracity of evidences, this paper can use transformer to learn the correlations of evidences and share it with the fake news detection task.

3) Evidence Selection and Representation: After obtaining the weighted evidence representations from the auxiliary task, this paper uses attention mechanism to select important evidences related to the claim. Given the claim representation R_c and evidence representation $\{h_1, h_2, \ldots, h_m\}$, this paper concatenates claim representation with each evidence representation:

$$a_j = R_c \oplus h_j \tag{5}$$

where h_j is the *j*-th evidence representation corresponding to claim.

This paper performs a linear and a softmax to calculate the attention score between the claim and the *j*-th evidence, and gains the evidence weighted representation based on claim-evidence attention R_e:

$$\alpha_j = \frac{exp(a_j W^T + b)}{\sum_j exp(a_j W^T + b)} \tag{6}$$

$$R_e = [\alpha_1 \cdot h_1, \ldots, \alpha_j \cdot h_j, \ldots, \alpha_m \cdot h_m] \tag{7}$$

where W^T denotes the weight matrix and b is the bias term, α_j is the attention score between the *j*-th evidence and the claim.

D. Fake News Classification

Given the visual representation and the textual representation, this paper uses a co-attention block to fuse the image representation R_v and image-related evidence representation R_e and obtains R'. The structure of the co-attention block is as follows:

$$R = R_e + MHA(R_e, R_v, R_v) \tag{8}$$

$$R' = R + FFN(R) \tag{9}$$

This paper feeds R' into a MLP layer to predict whether the news post is fake or not. The loss function of this part $\mathcal{L}_n$ is as follows:

$$\mathcal{L}_n = -\sum_{i}^{n} [y_i * log(\hat{y}_i) + (1 - y_i) * log(1 - \hat{y}_i)] \tag{10}$$

where y_i denotes the ground-truth label of post i and $\hat{y}_i$ indicates the predicted probability of being fake news.

The overall objective function consists of two parts: evidence classification loss and news classification loss. According to Equations 4 and 10, the objective function of MML can be defined as:

$$\mathcal{L}_{final} = \lambda \mathcal{L}_e + \mathcal{L}_n \tag{11}$$

where $\mathcal{L}_e$ denotes the evidence classification loss, $\mathcal{L}_n$ represents the news classification loss, and λ is a hyperparameter used to balance these two losses.

IV. Experimental Results and Discussion

In this section, authors conduct experiments to evaluate the effectiveness of the proposed MML. Specifically, the section aims to answer the following research questions: EQ1: Can MML improve the performance of multi-modal fake news detection? EQ2: Are visual representation and multi-task learning useful for fake news detection task? If so, how much can it improve? EQ3: Is MML model sensitive to different parameter settings?

A. Dataset

Authors evaluate the proposed MML model on the CCMR dataset, which is a multimedia fake news verification dataset with 17 events in total [21]. The dataset consists of 15,629 tweets with multimedia information, 4,625 webpages from Google and 2,506 webpages from Baidu that share similar multimedia content. Among them, this paper only uses the tweets and webpages from Google in the dataset to perform the experiment. Table I shows the statistical information of the CCMR dataset.

B. Baseline Methods

This section compares the proposed MML model with the following state-of-the-art methods:

1) **SpotFake+:** Singhal et al. [28] build a multi-modal fake news detection method based on transfer learning. The model extracts the features of text and image respectively, and then feeds the feature vectors to a fully connected layer for classification.

2) **IDM-FND:** Singhal et al. [29] develop a fake news detection framework based on inter-modality inconsistency. Firstly, the framework captures the relationship (inconsistency) among various components in news articles. Then, the features of text and image features are extracted and concatenated to detect fake news.

3) **MVNN:** Qi et al. [26] propose a multi-domain visual neural network framework, which extracts and fuses the features of frequency domain and pixel domain of images to detect fake news.

4) **MCAN:** Wu et al. [11] propose a multi-modal co-attention network to fuse the features of textual and visual features. Firstly, the network uses BERT to extract features. Secondly, the spatial domain and frequency domain features of the image are captured respectively. Finally, the multi-modal features are fused by stacking four co-attention layers.

5) **TFG:** Wen et al. [21] use cosine similarity and agreement classifiers to obtain the classification features. The network leverages the multimedia information to find the consistency and inconsistency among news from different social media platforms but sharing similar visual contents.

TABLE I. Statistics of the CCMR Dataset

ID	Event	Twitter	Google
01	Hurricane Sandy	10222	2204
02	Boston Marathon bombing	533	722
03	Sochi Olympics	274	347
04	MA flight 370	310	323
05	Bring Back Our Girls	131	108
06	Columbian Chemicals	185	63
07	Passport hoax	44	26
08	Rock Elephant	13	20
09	Underwater bedroom	113	59
10	Livr mobile app	9	15
11	Pig fish	14	20
12	Solar Eclipse	277	143
13	Girl with Samurai boots	218	60
14	Nepal Earthquake	1360	424
15	Garissa Attack	79	63
16	Syrian boy	1786	8
17	Varoufakis and zdf	61	20
	Total	15629	4625

C. Evaluating Metrics

To evaluate the performance of the proposed MML model, this paper uses four commonly used evaluation metrics: Accuracy, Precision, Recall and F1-score. Accuracy is a relatively intuitive evaluation index, which indicates the proportion of correctly classified samples in the total number of samples. Precision (P) represents the probability that the samples predicted to be true are real positive samples. Recall (R) represents the probability that positive examples in the sample are predicted to be correct. In practical evaluation of a model, both Precision and Recall should be considered, but it is difficult to compare the two values in a balanced way. The F1-score (F1) is a common method of integrating two values for evaluation:

$$F1 = \frac{2 * P * R}{P + R} \tag{12}$$

D. Implementation Details

This paper uses event 1-11 for training and event 12-17 for testing according to Wen et al. [21]. This paper sets the number of hidden layers in the Transformer encoder and the number of attention heads to 12. The maximum sequence length is set to 512. The learning rate is set to 1e-5 and the batch size is set to 8. The dropout of each layer is 0.1. The hyperparameter λ is 0.2.

E. Experimental Results and Analysis

This section compares the performance of the proposed model MML with the above baselines. From the results in Table II, authors can draw the following conclusions:

TABLE II. Results of MML Model and Baseline Models

Methods	Accuracy	Precision	Recall	F1-score
SpotFake+	0.7615	0.8212	0.7652	0.7921
IDM-FND	0.7937	0.7849	0.8231	0.8035
MVNN	0.8399	0.8173	0.8461	0.8315
MCAN	0.8573	0.8632	0.8347	0.8487
TFG	0.8912	0.8813	0.9254	0.9029
MML	**0.9225**	**0.9169**	**0.9262**	**0.9215**

TABLE III. Evaluation Results of the MML Model and Two Variants

Methods	Accuracy	Precision	Recall	F1-score
MML	**0.9225**	**0.9169**	**0.9262**	**0.9215**
MML-*w/o* Visual Representation	0.8501	0.8742	0.8439	0.8588
MML-*w/o* Evidence Veracity Classification	0.8651	0.8673	0.8426	0.8548

1) The MML model performs significantly better than all baseline models, achieving the Accuracy of 0.9225 and F1-score of 0.9215. Compared with SpotFake+ and IDM-FND, which simply combine the multi-modal features, MML achieves the greatest improvement, 16.1% in Accuracy and 12.9% in F1-score.

2) Compared with other multi-modal models MVNN and MCAN, the proposed MML model improves the Accuracy by 8.2% and 6.5%, respectively. It can be speculated that external evidence can effectively identify the correlations between text and image, and help improve fake news detection.

3) Compared with TFG, which also uses external evidence to detect fake news, MML is 3.1% and 1.8% higher in Accuracy and F1-score, respectively. This is because MML has advantages in extracting physical features of images and selecting important evidences.

F. Ablation Experiment

In this section, authors discuss the contribution of different components in the model, including visual representation learning and evidence veracity classification task. Authors remove the above two modules from MML model to obtain the following two variants: **MML- w/o Visual Representation**, which denotes MML only models the textual representation, and **MML- w/o Evidence Veracity Classification**, representing MML without multi-task learning.

The results of the two variants are shown in Table III. When the visual representation learning module is removed, the Accuracy and F1-scores drop to 0.8501 and 0.8588, respectively, showing the importance of visual representation for multi-modal fake news detection. By comparing MML with the

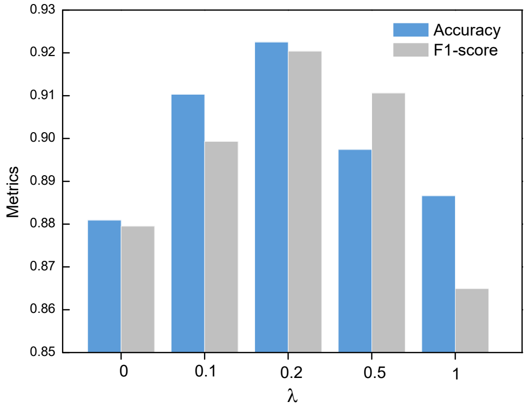


Fig. 3. Results of different Hyperparameter λ.

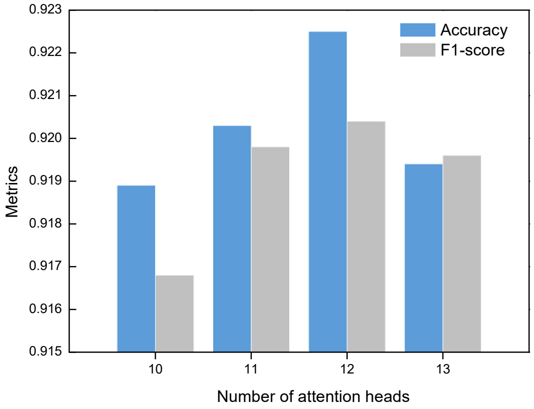


Fig. 4. Results of different Number of Attention Heads.

variant without evidence veracity classification task, authors can observe that all the evaluating metrics decreased greatly, which demonstrates the effectiveness of multi-task learning and the necessity of evidence selection and representation.

G. Parameter Sensitivity Analysis

To analyze the influence of hyperparameters on model performance, authors conduct the following two parameter sensitivity experiments.

1) Effects of hyperparameter λ: Note that λ is a weight parameter for balancing the evidence classification loss $\mathcal{L}_e$ and the news classification loss $\mathcal{L}_n$. In other words, the larger λ is, the greater the effect of evidence weight learning on fake news detection. This paper sets λ to 1, 0.5, 0.2, 0.1 and 0. The Accuracy and F1-score of different hyperparameter λ are shown in Fig. 3. Authors find that the model achieves the best performance when λ is 0.2, while the evidence classification loss brings an improvement of 5% on F1-score for the proposed MML model (λ=0, without evidence classification loss). This proves the effectiveness of the model by introducing evidence classification loss.

2) Number of attention heads: As shown in Fig. 4, authors can clearly see that the performance of the proposed model varies with the number of attention heads (i.e. 10, 11, 12 and 13). With the increase of the number of attention heads, the Accuracy and F1-score firstly increase and then decrease, and the best effect is achieved when the number of heads is 12.

V. Conclusion and Future Enhancements

This paper proposes a Multi-model Multi-task Learning model (MML) to detect multi-modal fake news on social media by modeling the image, claim and image-related evidence. By comparing MML with other competitive baseline methods, authors find that it is effective to use external evidence in this task, with an accuracy of 92.2%. In addition, besides the news classification loss, MML also introduces evidence classification loss to further optimize the model performance. By testing MML with different settings, authors observe that the proper setting of evidence classification loss can improve the performance of fake news detection. Finally, the results of the ablation experiments show that visual feature representation and evidence representation learning are beneficial to improve the fake news detection results, and the model is improved by 7.2% and 5.7%, respectively.

In the future, the authors are willing to extract the visual entity of the image and the text embedded in the image, and model them with the news text to further capture the correlation between the image and the text in multi-modal news.

Acknowledgment

Authors thank the anonymous reviewers for their constructive suggestions. This research was funded by the National Natural Science Foundation of China [61806049] and the Heilongjiang Postdoctoral Science Foundation [LBH-Z20104]. Corresponding author: Yang Li, yli@nefu.edu.cn.

References

[1] S. Kogan, T. J. Moskowitz, and M. Niessner, “Fake news: Evidence from financial markets,” *Available at SSRN*, vol. 3237763, 2019, doi: 10.2139/ssrn.3237763.

[2] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–36, 2017, doi: 10.1257/jep.31.2.211.

[3] A. Depoux, S. Martin, E. Karafillakis, R. Preet, A. Wilder-Smith, and H. Larson, “The pandemic of social media panic travels faster than the covid-19 outbreak,” p. taaa031, 2020, doi: 10.1093/jtm/taaa031.

[4] Q. Sheng, X. Zhang, J. Cao, and L. Zhong, “Integrating pattern- and fact-based fake news detection via model preference learning,” in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 1640–1650, doi: 10.1145/3459637.3482440.

[5] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, “Novel visual and statistical image features for microblogs news verification,” *IEEE transactions on multimedia*, vol. 19, no. 3, pp. 598–608, 2016, doi: 10.1109/TMM.2016.2617078.

[6] C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on twitter,” in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 675–684, doi: 10.1145/1963405.1963500.

[7] Y. Fung, C. Thomas, R. G. Reddy, S. Polisetty, H. Ji, S.-F. Chang, K. McKeown, M. Bansal, and A. Sil, "Infosurgeon: Cross-media fine-grained information consistency checking for fake news detection," in *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)*, 2021, pp. 1683–1698, doi: 10.18653/v1/2021.acl-long.133.

[8] Y. Wang, F. Ma, H. Wang, K. Jha, and J. Gao, "Multimodal emergent fake news detection via meta neural process networks," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 3708–3716, doi: 10.1145/3447548.3467153.

[9] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "Eann: Event adversarial neural networks for multi-modal fake news detection," in *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, 2018, pp. 849–857, doi: 10.1145/3219819.3219903.

[10] V. K. Singh, I. Ghosh, and D. Sonagara, "Detecting fake news stories via multimodal analysis," *Journal of the Association for Information Science and Technology*, vol. 72, no. 1, pp. 3–17, 2021, doi: 10.1002/asi.24359.

[11] Y. Wu, P. Zhan, Y. Zhang, L. Wang, and Z. Xu, "Multimodal fusion with co-attention networks for fake news detection," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2021, pp. 2560–2569. Available: https://aclanthology.org/2021.findings-acl.226.pdf.

[12] P. Qi, J. Cao, X. Li, H. Liu, Q. Sheng, X. Mi, Q. He, Y. Lv, C. Guo, and Y. Yu, "Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 1212–1220, doi: 10.1145/3474085.3481548.

[13] Z. Liu, C. Xiong, M. Sun, and Z. Liu, "Fine-grained fact verification with kernel graph attention network," *arXiv preprint arXiv:1910.09796*, 2019, doi: 10.48550/arXiv.1910.09796.

[14] W. Zhong, J. Xu, D. Tang, Z. Xu, N. Duan, M. Zhou, J. Wang, and J. Yin, "Reasoning over semantic-level graph for fact checking," *arXiv preprint arXiv:1909.03745*, 2019, doi: 10.48550/arXiv.1909.03745.

[15] J. Ma and W. Gao, "Debunking rumors on twitter with tree transformer." ACL, 2020. Available: https://ink.library.smu.edu.sg/sis_research/5599.

[16] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, and J. Huang, "Rumor detection on social media with bi-directional graph convolutional networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 549–556, doi: 10.1609/aaai.v34i01.5393.

[17] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "defend: Explainable fake news detection," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 395–405, doi: 10.1145/3292500.3330935.

[18] L. Wu, Y. Rao, Y. Lan, L. Sun, and Z. Qi, "Unified dual-view cognitive model for interpretable claim verification," *arXiv preprint arXiv:2105.09567*, 2021, doi: 10.48550/arXiv.2105.09567.

[19] N. Vo and K. Lee, "Where are the facts? searching for fact-checked information to alleviate the spread of fake news," in *The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, 2020, doi: 10.48550/arXiv.2010.03159.

[20] Z. Jin, J. Cao, Y. Zhang, and Y. Zhang, "Mcg-ict at mediaeval 2015: Verifying multimedia use with a two-level classification model." in *MediaEval*, 2015. Available: http://ceur-ws.org/Vol-1436/Paper51.pdf.

[21] W. Wen, S. Su, and Z. Yu, "Cross-lingual cross-platform rumor verification pivoting on multimedia content," *arXiv preprint arXiv:1808.04911*, 2018, doi: 10.48550/arXiv.1808.04911.

[22] E. Kochkina, M. Liakata, and A. Zubiaga, "All-in-one: Multi-task learning for rumour verification," *arXiv preprint arXiv:1806.03713*, 2018, doi: 10.48550/arXiv.1806.03713.

[23] J. Ma, W. Gao, and K.-F. Wong, "Detect rumor and stance jointly by neural multi-task learning," in *Companion proceedings of the the web conference 2018*, 2018, pp. 585–593, doi: 10.1145/3184558.3188729.

[24] L. Wu, Y. Rao, H. Jin, A. Nazir, and L. Sun, "Different absorption from the same sharing: Sifted multi-task learning for fake news detection," *arXiv preprint arXiv:1909.01720*, 2019, doi: 10.48550/arXiv.1909.01720.

[25] Q. Li, Q. Zhang, and L. Si, "Rumor detection by exploiting user credibility information, attention and multi-task learning," in *Proceedings of the 57th annual meeting of the association for computational linguistics*, 2019, pp. 1173–1179, doi: 10.18653/v1/P19-1113.

[26] P. Qi, J. Cao, T. Yang, J. Guo, and J. Li, "Exploiting multi-domain visual information for fake news detection," in *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2019, pp. 518–527, doi: 10.1109/ICDM.2019.00062.

[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018, doi: 10.48550/arXiv.1810.04805.

[28] S. Singhal, A. Kabra, M. Sharma, R. R. Shah, T. Chakraborty, and P. Kumaraguru, "Spotfake+: A multimodal framework for fake news detection via transfer learning (student abstract)," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 10, 2020, pp. 13 915–13 916, doi: 10.1609/aaai.v34i10.7230.

[29] S. Singhal, M. Dhawan, R. R. Shah, and P. Kumaraguru, "Inter-modality discordance for multimodal fake news detection," in *ACM Multimedia Asia*, 2021, pp. 1–7, doi: 10.1145/3469877.3490614.

A Model Driven Approach for Unifying User Interfaces Development

Henoc Soude
Institut de Mathématiques et des
Sciences Physiques, BENIN

Kefil Koussonda
Epitech, FRANCE

***Abstract*—In this paper, we dealt with the rapid development of client web applications (frontend) in a context where development frameworks are legion. In effect with the digital transformation due to the COVID-19 pandemic we are witnessing an ever-increasing demand of the application development in a relatively short time. To this is added the lack of skilled developers on constantly evolving technologies. We therefore offer a low-code platform for the automatic generation of client web applications, regardless of the platform or framework chosen. First, we defined an interface design methodology based on a portal. We then implemented our model driven architecture which consisted of defining a modeling and templating language, centered on user data, flexible enough to not only be used in various fields but also be easily used by a citizen developer.**

Keywords*—*Model driven; user interface; modeling language; templating language; low code; citizen developer

I. Introduction

Over the past two decades, web applications have evolved and are now used in all areas of everyday life: medical, transport, e-commerce and social networks. One of the reasons for their success is that they can be used on all our devices like laptops, smart phones, tablets and desktops. The interfaces are determining elements in the acceptance of a web application [1], [2]; their development is therefore of paramount importance hence the multiplicity of frameworks in the market. For some years the frameworks *javascript* have spread widely to the point of becoming indispensable in the development of modern applications. By the way Sacha Greif et al. [3] have launched an annual survey since 2016 to analyze developments and trends in *javascript* frameworks by focusing on based data from more than 20,000 developers spread over more than 100 country. This study shows us, among other things, the most used frameworks (react, angular, vue) but doesn't give us any information on how they choose, knowing that they have different learning curves more or less complex. Indeed, the COVID-19 pandemic requires us to rapidly develop more and more complex applications, in the context of a global shortage of developers [4], [5].

Model-based development (MDD) has been used in the literature [6]–[8] in order to simplify the development of applications and increase productivity. It is a development paradigm which according to Stephen et al. [9], consists in building the model of a system and then generate a real instance of this system from the obtained model. MDD was quickly adopted by the Object Management Group (OMG) which standardized model-based architecture (MDA) and is composed of several levels, two of which are essential in our development context of web application: platform independent model (PIM); used to describe the system and platform specific model (PSM); used to generate an instance actual system. The PIM and PSM models are described using languages that are either standardized or domain-specific (DSL). Most of these languages are neither suitable, nor flexible enough and too complex [10], in terms of learning curve.

One of the solutions to address the problems related to developer shortage is the use of low-code platforms [11], which in addition to using model-based approaches, use graphical or high-level languages to make development accessible to all. According to a Gartner report more than 50% of companies will adopt them as a solution strategic by 2023 [12]. There are many low code [13]–[15] which agrees with generating interfaces from models. They have the particularity of either generating code for their own platform or to use languages specific to their domain wich most of the time are difficult to handle. Added to this is the fact that the adoption of low-code platforms by citizen developers is conditioned by the simplicity and accessibility of the underlying models.

This work is part of a more global project to implement a low code [16] platform to automate the development of tasks in various fields. For the sake of consistency, all projects share not just the same development languages: *c/c++* and exchange data: *json*, but also the same automation system; the generator must generate codes in several domains without requiring the intervention of expert developers.

The general research question of our work is: *How can we unify user interface (UI) development?*. Whatever framework used: Mithril, React, Vue, the development process must be identical and the learning curve should be as simple as possible even for a person without any developing skills (citizen developer). The solution we propose is to automatically generate the code for the different platforms from a single model, defined by the user. It comes in the form of the following contributions:

- the development of a portal and the definition of an UI design methodology based on a portal. It allows to gain in productivity and to make the implementation of UI simpler since their navigation elements have to be configured.
- the development of a cloud-based graphic editor for UI modeling
- the definiton of a modeling language which has simple structure; all its elements have the same basic,

and flexible structure; it can be extended for new platforms.

- the development of a code generation system consisting of an engine and a template language.

The rest of the document is organized as follows: Section II presents state of the art. Then we described our model-based approach to Section III. Our interface modeling language and our template language are described in Sections IV and V. In Section VI we presented a general discussion of work then we concluded with the Section VII

II. Related Work

Whatever platform or model-based approach used, they are distinguished by the modeling language and the generation system used.

A. UI Modeling Language

The principle of modeling a user interface consists in finding an abstract representation of the constituent elements of the interface. This representation usually contains information relating to the organization, the description of the elements in the interface and interactions with users. This explains the multiplicity of modeling languages that differ depending on the type and storage of the information they contain.

UIML [17] is the first universal modeling language independent of the platforms and technologies used. It is an XML dialect which describes an interface in five sections: the description, the structure, the data, the style, and the event. Subsequently the language was standardized by OASIS group. USIXML [18] is another language based on XML and standardized by the World Wide Web Consortium (W3C) whose particularity is to describe the multimodal interfaces. The proposed language in our work differs from the two languages presented by the format of the definition of models; we use the *json* format unlike the XML used in most templates [19], [20].

Brambilla *et al.* [21] present interaction flow modeling language (IFML) which was quickly adopted as a standard by OMG. Language allows to describe the structure and interactions with end users via a set of visual components representing the different elements of an interface. This concept of visual modeling has greatly appealed whether in the academic world or that of industry [22]–[26]. The language has introduced a high level of abstraction that makes an element of the model we can match several components of an interface. This ambiguity is resolved by the language when generating the interface, since the user provides these correspondence elements. This constitutes a real problem in the works as shown where we have no idea of the interface that the end user wants to generate: we cannot provide a mapping that will satisfy all users.

More recently Moldovan *et al.* [27] presented a model-based approach for developing user interfaces f' or multi-target applications. They first define their modeling language (OpenUIDL) whose syntax is based on the JSON format. Then they present the different stages of their approach for the development of interfaces: the design, the generation and the deployment of the interface. Although dealing with the same problem with an almost identical approach, our work differs mainly in the level of the semantics of the modeling language and the process of generation. In order to improve the problem of limited accessiblity to defintions of modeling languages (cf. [27]), we propose a model of a simpler and more intuitive level than the authors of [27]. Indeed in their model a node can be of type static value, dynamic reference, element, conditional, repeat, slot, nested-styled and its content is specified through the attribute of the same name. So you need at least two objects *json* nested to define an element of an interface then in our model the equivalent of the node directly represents an element of the interface. Their code generation model remains limited to web interfaces and cannot allow ”citizen developers” to generate code or data in specific areas contrary to what we propose.

B. Code Generation

In the context of automatic web interface generation, two approaches are often used: that based on the structures tree abstracts [27] and that of the engines template. The major drawback of the first approach is that it requires having technical skills if you want to generate code in a language other than the original one. We are indeed in need of a system that allows seasoned users or not to generate code or documents in various fields.

The second approach is more appropriate in our context since the engines offer a template description language; the user will only have to describe his new model through this language. Most languages used for the development of java web interfaces (freemaker [28]), python (jinja2 [29]) and javascript(mustache [30]) all have template engines whose languages does not tell us nor inhibit us from defining the models of views whose structure is not known during design. XSLT [31] is a language standardized by W3C which allows you to transform XML documents into another format (HTML, XML, js,etc.). It easily solves the problem mentioned above thanks to these directives*template* and *apply*; unfortunately there remains a verbose language and very complex to handle.

III. Model based Approach

The objective of our work is to allow developers to implement web applications without worrying about platforms or frameworks available in the market. For this we propose, as indicated in the Fig. 1 a three-step approach: design, build, and refinement of the interface after deployment. Except for the refinement, the other two steps are MDD classics. The distinctiveness of our solution lies in the methodology, the concepts, the languages and tools we use in the different stages.

A. UI Design

The first principle of our design approach is that the user doesn't waste time implementing common and recurring concepts in web applications: menu navigation and operation application on user data. For this we have implemented an application portal which not only is an application receptacle but also has mechanisms for specifying menu and application-related actions. As shown, Figure 2, is divided into five parts: (i) the main bar which indicates the name of the current application and different menus drop-down such as the list

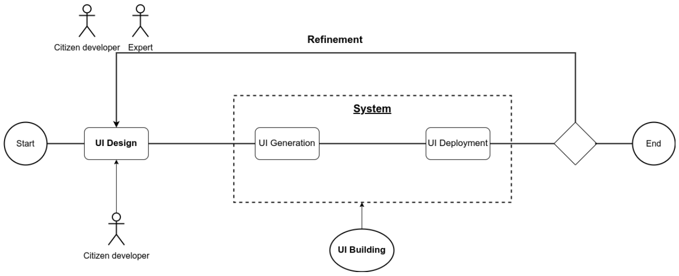


Fig. 1. The Different Steps of our Model-based Approach.

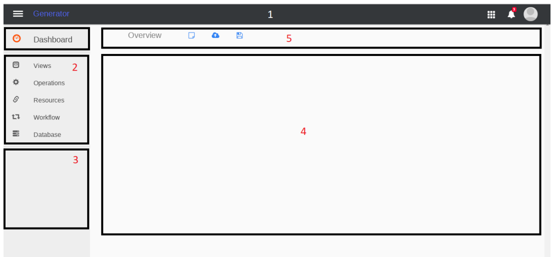


Fig. 2. The Structure of our Application Portal.

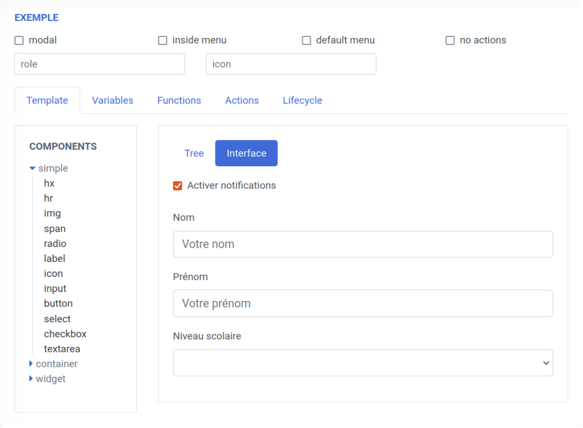
of applications, (ii) represents the elements of menu of the current application, (iii) represents the list of applications related to the current application, (iv) represents the workspace in which the different views of the application will be hosted and (v) represents the action bar which indicates the possible operations on a given view. The design of an application's interfaces therefore consists of identifying and declaring the set of views or interfaces that compose it and for each of it to define its content and its operations.

The second principle of our approach is that of the separation of concerns that we have implemented in our editor interface graphic: users of different level techniques will be able to intervene on different parts as shown in Fig. 3 the design is divided into six sections: *view*, *template*, *variables*, *functions*, *actions* and *lifecycle*.

The *view* section describes the organization of a view in the application portal. The form elements `inside menu`, `role` and `icon` allow resp. to specify whether the view is accessible or not in the application menu, the role needed to access the view and view icon in the menu. The `modal` attribute indicates that the view can be imported by another view.

The *template* section describes the contents of a view. The user uses the different predefined elements to describe the content of its view. The `tag` can be used to specify non-predefined elements. Every Adding of an element, the user can choose either to see the preview (tab interface) or to see the generated model (tree tab). At any moment of the design the user can select an element of the model and modify its properties.

The *variables* section describes the variables used by the different view elements. They are usually created by the user which can also associate test values to then. The *functions* section describes user-defined methods. The *actions* section describes sight operations. The user must provide the icon as well as the method to call. Finally the *lifecycle* section describes the mechanisms of *javascript* component status notification.



EXEMPLE

□ modal □ inside menu □ default menu □ no actions

role

icon

Template Variables Functions Actions Lifecycle

COMPONENTS

- simple
  - hx
  - hr
  - img
  - span
  - radio
  - label
  - icon
  - input
  - button
  - select
  - checkbox
  - textarea
- container
- widget

Tree Interface

☑ Activer notifications

Nom

Votre nom

Prénom

Votre prénom

Niveau scolaire

Fig. 3. The Interface Confuguration of a View.

B. UI Building

The user does not intervene in the process of generating the views even if he is the instigator. As shown in Fig. 4, our template takes as parameter the view model and the mapping of the elements of the seen. The mapping of an element corresponds to its html template for a support given. We have therefore provided the mapping of all the predefined elements of our editor for the following frameworks: mithril.js, react.js, vue.js. An example of mapping is presented in Section V-B.

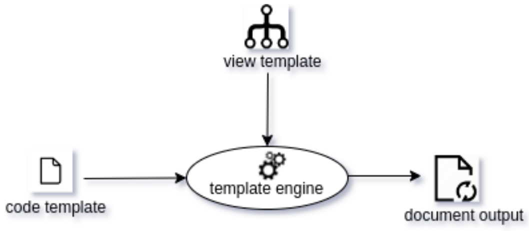


Fig. 4. Our Code Generation Process.

In addition to specificities related to the generation of web applications our build system needs to be flexible enough to not only allow generation in multiple domains but also allow citizen developers to edit their own model for specific needs. For this we have proposed a model language based on text and simple substitution directives, intuitive and non-verbose. We have also decided to limit the number of instructions to make it easier to get started. We have implemented a simple and adaptable algorithm to all situations. The algorithm consists of traversing a tree structure (in *json* format) and each node of the tree generates the corresponding code by retrieving the associated template to its type. This implies that any *json* object passed as a parameter must have the `type` attribute.

C. UI Refinement

After the generation of the interfaces, the user can deploy the application on our beta server to be tested by all connected users. This mechanism allows us to integrate the Agile principles in our view generation process: the end users or team members can make feedback that will be integrated during this phase.

The positive point of our refinement mechanism associated with principle of the separation of concerns developed during the phase of design is that it is possible for a citizen developer to design the organizational aspect of the views: positioning of the different interface elements. Once his work has been validated then a user with technical skills will be able to finalize the views during the refinement phase.

IV. UI Modeling Language

The most natural representation of an interface is that of a tree in which a node corresponds to an element of the interface. The Fig. 5 is a representation in the form of a tree of the html component `table`: it has `tr` elements which, in turn, have `td` elements. Our modeling therefore consists in defining in a format *json* the structure of interfaces. The nodes of the description tree are represented by the objects*json* whose structures are defined via their json schema.

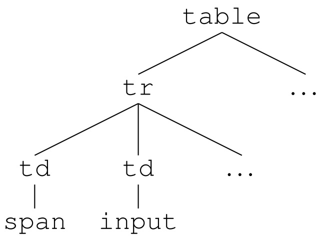


Fig. 5. Representation of `table` Component as a Tree.

A. Syntax of Models

Fig. 6 presents the schema *json* of the basic structure of elements of an interface. An element is described by its type and by either its value or its descendants. All other properties of the element are optional; our generator is able to generate values for them by default. The `type` attribute is a character string whose value is the name of the element with which the object is associated. The `children` attribute is an array that describes the children of an element. The attribute `value` is a character string containing the value of an element. The semantics of the value of an element depending on its type. For example the value of a `button` is the text that accompanies it while the value of a `span` represents its content.

```
{
  "$id": "schema/base",
  "type": "object",
  "properties": {
    "type":    {"type": "string"},
    "value":   {"type": "string"},
    "children":{"type": "array",
    "items":   {"$ref": "#"}}
  },
   "required": ["type",
     {"oneOf": ["value", "children"]}]
 }
```

Fig. 6. The Basic Structure of an Element of our Model.

The value of an element can either be literal or come from a variable. For this reason we use a formatting which consists of preceding the value of the attribute `value` by one of the following characters: `$, #, %` . When the attribute value is not prefixed for any of these characters then the value is used as such. When preceded by the character`#` then it should be treated as given *json*. When she is preceded by the character `$` then the value of the element comes from a variable, defined inside the html component, whose name is the value of the attribute without the prefix character. When the attribute value is preceded by the character `%` then the value of the element comes from a variable passed as a parameter to the html component. In the following example {``a": ``foobar", ``b": ``$foo", ``c": ``%bar", ``d": ``#{}"}, a, b, c, d have the respective value `foobar`, the value of the variable`foo`, the variable of the variable `bar` passed as component parameter and an empty json object.

```
{
    "$id": "schema/view",
    "type": "object",
    "properties": {
        "type": {"type": "string"},
        "imports":  {"type": "array",
            "items": {"$ref": "#/$defs/keyval"}},
        "variables":{"type": "array",
            "items": {"$ref": "#/$defs/keyval"}},
         "functions":{"type": "array",
             "items": {"$ref": "#/$defs/keyval"}},
         "template": {"type": "object",
             "properties": {"$ref": "schema/base"}},
         "oninit":   {"type": "string"},
         "oncreate": {"type": "string"},
         "onupdate": {"type": "string"},
         "onremove": {"type": "string"}
     },
     "required": ["type", "variables",
                 "modals", "functions", "template"],
     "$defs": {"keyval":{
         "type": "object",
         "properties": {
           "name"  {"type":"string"},
           "value" {"type":"string"}
         }
     }}
 }
```

Fig. 7. Metamodel of the View Component.

```
{
    "$id": "schema/input",
    "type": "object",
    "properties": {
      "type":       {"type": "string"},
      "value": {"type": "string"},
      "subtype": {"enum":["text", "password",
                 "date", "email", "number", "file"]}
      "label":   {"type": "string"},
       "placeholder": {"type": "string"},
       "class":   {"type": "string"},
       "style": {"type": "string"},
       "onchange": {"type": "string"}
     },
     "required": ["type", "subtype", "value"],
     "additionalProperties": true
 }
```

Fig. 8. Metamodel of Input Component.

```
{
    "$id": "schema/tag",
    "type": "object",
    "properties": {
      "type":  {"type": "string"},
      "value": {"type": "string"},
      "class": {"type": "string"},
      "style": {"type": "string"},
      "children": {"type": "array",
          "items": {"$ref": "schema/base"}},
     },
     "required": ["type", "value", "children"]
 }
```

Fig. 9. Metamodel of the Tag Component.

```
{
    "$id": "schema/list",
    "type": "object",
    "properties": {
      "type": {"type": "string"},
      "data": {"type": "string"},
      "root": {"type": "string"},
      "class":{"type": "string"},
      "style":{"type": "string"},
       "iterator": {"type": "string"},
       "children": {"type": "array",
          "items": {"$ref": "schema/base"}},
     },
     "required": ["type", "data", "children"]
 }
```

Fig. 10. Metamodel of List Component.

```
{
    "$id": "schema/wizard",
    "type": "object",
    "properties": {
      "type": {"type": "string"},
      "done": {"type": "string"},
      "data": {
        "type": "array",
        "items": {"$ref": "#/$defs/pagedef"}},
     },
     "required": ["data"],
     "$defs": {"pagedef":{
         "type": "object",
         "properties": {
           "name" {"type":"string"},
           "content" {"type":"string"}
         }
     }}
 }
```

Fig. 11. Metamodel of Wizard Component.

B. Components

We have predefined a number of interface elements which we have organized into three categories:

- *simple*; they have no child elements. hr, hx, img, span,radio, label, icon, input, button, select, checkbox, textarea.
- *container*; they have child elements. nav, tab, link, list, form, group, table, accordion, dropdown, paragraph.
- *widgets*; these are non-standard html elements; wizard, treeview, carousel, display.

We have not implemented all the elements of an interface but rather those that come up regularly in web applications. Nevertheless the tag element can be used to implement elements not predefined.

In the rest of the document we present only the special components (view and tag) and one component per category due to the simplicity of the model.

a) view: An element of type view is a special element allowing to model the organization of interfaces in the application portal (see section III-A). As shown in its json schema, in Figure 7, the attributes variables, imports, functions and template are mandatory. The template attribute is an object representing the content of sight. The attributes variables, imports and functions represent respectively the variables, the imported elements and the user-declared functions.

b) input: Fig. 8 represents the model definition of a input. The attributes value and subtype are required. This model represents multiple types of input via the subtype attribute; other attributes are added to the model depending on the type chosen. The label and the placeholder can also be specified via attributes of the same name. The user can specify the action to perform when changing the value via the onchange attribute.

c) tag: The tag type element is used to model dynamic attribute (unknown at design time) or non-predefined elements. As shown in Fig. 9 the attributes value and children are mandatory. The children attribute contains the definition of the element's children given while the attribute value represents the real no of the element. The element name can be assigned literally or via a variable (cf. Section IV-A).

d) list: The list type element is used to model components from a repetitive action; this is the case for example of the elements dl, ul and table. As shown in Fig. 10 the attributes data and children are required. children contains the definition of the elements that are repeated according to the data coming from the data attribute which must be an array. The root attribute when present becomes the parent element containing the repeated elements. The iterator attribute represents the name of the iterator to use in the code for traversing data in data.

e) wizard: Fig. 11 represents the model definition of a wizard. The attribute data is mandatory and represents all the pages of the wizard. Pages are defined by an object whose attribute name represents the page name and the content attribute represents the content of the page. The content of a page is a reference that is resolved when generated. The action to be performed after the process is completed is specified via the done attribute.

V. Template Language

The language that we propose must not only make it possible to write the model codes of our web interfaces but also be sufficiently simple and flexible to allow a citizen developer to write their own models in various areas. For this we have chosen a language composed of texts: they are copied as such by the template engine, and substitution directives: they are replaced by the value of the variables or expressions they represent.

In order to make the language simple and accessible, the transformation of directives is solely based on user data; unlike some languages [28], [31], ours does not allow for defining variables or functions in our models.

User data is passed as a parameter to the template engine as an object of *json* (see Fig. 12) whose attributes are the variables used by the substitution directives. The language uses the symbols `{,},[,]` as block delimiters, the character `%` to escape special characters and the character `$` to introduce directives.

```
{
  "string": "foobar", "number": 12.23,
  "bool": false, "null": null,
  "object": {"key": "val"},
  "array":  [10,20,30]
}
```

Fig. 12. An Example of User Data.

A. Directives

a) content: This directive is used to retrieve the value of an attribute of the data passed as a parameter to the engine.

$$\$ < attribut >< . < attribut_1 >< \cdots . < attribut_n >>> \tag{1}$$

$$\$ < attribut > [< index >] \tag{2}$$

Its syntax is presented by the expression 1 where $< attribute >$ represents the attribute whose value we are looking for. Considering the data of Fig. 12, the directives $\$string$, $\$number$, $\$bool$, $\$ull$, $\$object$ and $\$array$ respectively have the value foobar, 12.23, false, null, an empty string and an empty string. The directives $\$object$ and $\$array$ have the value of empty strings since we must specify the element that is sought within them. With regard to the directive $\$object$ we use the optional part of the expression 1 by preceding the name of the element with the character .. The $\$object.key$ directive is used to retrieve the value of the key. The expression notation 2 is used for arrays. The pattern $< index >$ indicates the position of the element of the array that we are looking for; in our example $\$array[2]$ and $\$array[\$number]$ have the value 20 and an empty string since $\$number$ is not an integer.

The evaluator of a directive returns the Boolean value false which indicates that an error occurred while evaluating and the value true in the opposite case. In case of error the evaluator generates an empty string.

b) alternative: This directive allows you to choose between two models. The expression 3 presents its syntax where $||$ represents the separator of the two patterns. The model $< model1 >$ is generated when it tests true while model $< model2 >$ is generated when it tests true when the first shows false. In considering the Figure data 12 the guidelines $\$string||\$number$ and $\$foo||nothing$ have the value foobar and nothing.

$$< model1 > || < model2 > \tag{3}$$

c) test: This directive makes it possible to choose between two models according to the value of a variable. The expressions 4 and 5 present its syntax.

$$\$ < attribut > \{< model1 >\}\{< model2 >\} \tag{4}$$

$$\$ < attribut >< CMP >< VAL > \{< model1 >\}\{< model2 >\} \tag{5}$$

In the expression 4 the test is implicit; the model $< model1 >$ is generated if attribute attribute is true else $< model2 >$ is generated. The attribute is considered true when it exists and value is neither false nor null. In the expression 5 we specify the comparator as well as the reference value. The comparator can be one of the operators following: $=, <, >, >=, <=, ! =$. Considering the data of Fig. 12 the directives $\$foo\{found\}\{not\ \ found\}$ and $\$number = 12.23\{equal\}\{not\ \ equal\}$ have value not found and equal.

d) repeat: This directive makes it possible to repeat the generation of the model according to the number elements in an array variable. The expression 6 presents its syntax where $< attribute >$ must be of type array.

$$\$ < attribut > *\{< model1 >\} \tag{6}$$

$$\$\$ < . < attribut >> \tag{7}$$

The expression 7 presents the iteration variable notation that represents an array element. When the element is an object then we use the optional part of the expression in order to specify the attribute to be to find the value. Considering the data in Figure 12 the directive $\$array * \{\$\$\}$ has the value 102030.

e) separator: This directive is used to automatically generate separators in the context of the *repeat* directive. The expression 8 presents its notation where $< SEPATOR >$ is the separator character. Considering the data of the Fig. 12 the directive $\$array * \{\$\$\$:\}$ has the value 10:20:30.

$$\$ < SEPATOR > \tag{8}$$

f) only if: This directive is used to generate a model if and only if all its directives evaluate to true. The expression 9 presents its notation. Considering the data in Fig. 12 the directives $[\$string\ has\ \$number]$ and $[\$name\ has\ \$foo]$ have for value foobar has 12.23 and an empty string.

$$[< model >] \tag{9}$$

g) built-ins: These are language variables whose notation is present by the expression 10 where $< func >$ represents the name of the variable and $< arg_i >$ represent the arguments associated with the variable. Arguments are optional and allow to change the behavior of the variable.

$$\$ < func ><, < arg_1 >, \cdot, < arg_n >> \$ \tag{10}$$

The variable `$index$` is used in a context of the directive *repeat* and returns the iteration number. The directive $\$array*$

$\{\$index\$\}$ has the value `012` if we consider the data of the Fig. 12.

The variable `$call$` is used in a context of the directive *repeat* and allows recursively calling the template engine on the elements of a table. Elements must be objects that have at least one attribute *type*.

The variable `$href$` is used when we want to retrieve data on media different from the object passed as a parameter to the engine. The data user must contain an *urn* attribute that describes the exact location of the data. The *urn* is composed of fragment separated by the character `:`. The first fragment can take the value `file`, `db`, `rest` depending on whether the data comes from a file, a database or a rest resource. The variable `$count$` is used in a context of the directive *repeat*; it returns the number of elements in an array. The directive $\$array * \{\$count\$\}$ has the value `3` if we consider the data of the Fig. 12.

The variable `$sum$` is used in a context of the directive *repeat*; it returns the sum of the elements in an array. The directive $\$array * \{\$sum\$\}$ has the value `60` if we consider the data of the Fig. 12.

h) Arithmetic operations: The language allows performing arithmetic operations only on the user data according to the syntax of the expression 11 where $< OP >$ Represents an operation among `+,-,/,*,%` and $< VAL >$ represents a whole value. Operations are only allowed for attributes of type *number* Or *string*. Only the addition operation is possible on strings; that is to reduce, from the beginning, the size of the string by the number $< VAL >$. The directives $\$string + 3$ and $name + 3$ have the value `bar` and `name+2`.

$$\$ < attribut >< OP >< VAL > \tag{11}$$

```
module.export= ()=>{
  $modals*{var $$.name%=$href$;}
  $variables*{var $$.name%=$$.value;}
  var $name%_cls = function(){
  return {
    $functions*{$$.name:$$.code$,}
  };}();
  return {
    oninit:(vnode)=>{
      app.setActions($noaction{null}{
        %[$actions*{$call$$,}%]
      });
      $oninit
    },
    [oncreate:(vnode)=>%{$oncreate%},]
    [onupdate:(vnode)=>%{$onupdate%},]
    view:(vnode)=>{
     return $template*{$call$};
    }
    [,onremove: (vnode)=>{$onremove}]
  };
}
```

Fig. 13. Mapping of `view` component for Mithril Framework.

```
<template>
  $template*{$call$};
</template>
<script>
$modals*{import {$$.name} from "$vref+5";}
export default {
  data() {
    return {
      $variables*{$$.name: $$.value$,}
    }
  },
  computed: {
    $functions*{$$.name$$.code$,}
  },
  beforeCreate(){
    app.setActions($noaction{null}{
      %[$actions*{$call$$,}%]
    });
    $oninit
  },
  [created(){$oncreate},]
  [updated(){$onupdate},]
  [destroyed(){$onremove}]
};
</script>
```

Fig. 14. Mapping of `view` Component for vue Framework.

B. Mapping

Mapping consists of using the language to describe the structure of the code to be generated for each element of our model. Here we present only the Mapping of the `view` element, as an example. Fig. 13 and 14 present the code structure of `view` For mithril and vue frameworks. Users can provide their mapping during the init phase of the engine.

VI. Discussion

How do we assess the learning curve of our modeling method Views? We compared our modeling approach to the one presented by Moldovan *et al.* [27] and that based on IFML. All the three approaches use graphic elements in order to simplify the Modelization. However, the other two approaches introduce new concepts, in addition to the components of an interface.

Why are we talking about unification when we have not limited ourselves on only three frameworks? We talk about unification because our system allows you to add other platforms or frameworks. Just provide the mapping (cf. section V-B) of the predefined elements for the said platform.

Is our built system as simple as claimed? We submitted, to fifteen students, the writing of templates in different fields using our language and XSLT. After analyzing their return from experience it appears that our work was preferred because of its syntax compact and its low learning curve.

Is our build system flexible enough to be used in different areas? Unlike the work of Moldovan *and Al.* [27] which use one generator per target, we use the same generator ourselves regardless of the target; only the mapping changes. Furthermore our generator was used to generate unit tests in some of our projects.

VII. Conclusion

In this work we presented a model-based approach for the development of web interfaces, whatever the platform and have implemented an application portal to simplify the Modeling work. All elements common to interfaces are implemented in the portal and configured during modeling. Then we implemented a cloud-based graphic editor for the modeling of interfaces. The elements of the interface are described via our model *json* whose flexibility allows implementing the abstraction of elements of different platforms without introducing new concepts. Finally we implemented an automatic code generation system for frameworks Mithril, React, Vue. The system is composed of a template language, with a low learning curve, and a template engine focused on user data: the generator determines the model to generate based on the type of the data. Their association make the generation system suitable for different areas.

To further reduce interface development time, we would like to generate them from their draft.

References

[1] E. Yigitbas, I. Jovanovikj, K. Biermeier, S. Sauer, and G. Engels, "Integrated model-driven development of self-adaptive user interfaces," *Software and Systems Modeling*, vol. 19, no. 5, pp. 1057–1081, 2020.

[2] M. Thomas, I. Mihaela, R. M. Andrianjaka, D. W. Germain, and I. Sorin, "Metamodel based approach to generate user interface mockup from uml class diagram," *Procedia Computer Science*, vol. 184, pp. 779–784, 2021.

[3] (2022) State of javascript. [Online]. Available: https://stateofjs.com

[4] Daxx. (2022) Us and the global tech talent shortage in 2022. [Online]. Available: httaps://www.daxx.com/blog/development-trends/software-developer-shortage-us

[5] T. Sloyan. (2021) Is there a developer shortage? yes, but the problem is more complicated than it looks. [Online]. Available: https://www.forbes.com/sites/forbestechcouncil/2021/06/08/

[6] S. Abrahão, E. Insfran, A. Sluÿters, and J. Vanderdonckt, "Model-based intelligent user interface adaptation: challenges and future directions," *Software and Systems Modeling*, vol. 20, no. 5, pp. 1335–1349, 2021.

[7] L. Alwakeel and K. Lano, "Model driven development of mobile applications," in *Doctoral Symposium, ECOOP 2020*, 2020.

[8] A. Sabraoui, A. Abouzahra, K. Afdel, and M. Machkour, "Mdd approach for mobile applications based on dsl," in *2019 International Conference of Computer Science and Renewable Energies (ICCSRE)*, 2019, pp. 1–6.

[9] W. B. Frakes and S. Isoda, "Success factors of systematic reuse," *IEEE Software*, vol. 20, no. 05, pp. 14–19, sep 1994.

[10] J. S. Mittapalli and M. P. Arthur, "Survey on template engines in java," in *ITM Web of Conferences*, vol. 37. EDP Sciences, 2021, p. 01007.

[11] A. C. Bock and U. Frank, "Low-code platform," *Business & Information Systems Engineering*, vol. 63, no. 6, pp. 733–740, 2021.

[12] P. Vincent, K. Iijima, M. Driver, J. Wong, and Y. Natis, "Magic quadrant for enterprise low-code application platforms," *Gartner report*, 2020.

[13] (2022) Appian. [Online]. Available: https://appian.com/platform/low-code-development/

[14] (2022) Mendix. [Online]. Available: https://www.mendix.com/

[15] H. Lourenço, C. Ferreira, and J. C. Seco, "Ostrich-a type-safe template language for low-code development," in *2021 ACM/IEEE 24th International Conference on Model Driven Engineering Languages and Systems (MODELS)*. IEEE, 2021, pp. 216–226.

[16] A. Sahay, A. Indamutsa, D. Di Ruscio, and A. Pierantonio, "Supporting the understanding and comparison of low-code development platforms," in *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 2020, pp. 171–178.

[17] M. Abrams, C. Phanouriou, A. L. Batongbacal, S. M. Williams, and J. E. Shuster, "Uiml: an appliance-independent xml user interface language," *Computer networks*, vol. 31, no. 11-16, pp. 1695–1708, 1999.

[18] J. Vanderdonckt, Q. Limbourg, B. Michotte, L. Bouillon, D. Trevisan, and M. Florins, "Usixml: a user interface description language for specifying multimodal user interfaces," in *Proceedings of W3C Workshop on Multimodal Interaction WMI*, vol. 2004. sn, 2004.

[19] S. Berti, F. Correani, F. Paterno, and C. Santoro, "The teresa xml language for the description of interactive systems at multiple abstraction levels," in *Proceedings workshop on developing user interfaces with XML: advances on user interface description languages*, 2004, pp. 103–110.

[20] F. Paterno', C. Santoro, and L. D. Spano, "Maria: A universal, declarative, multiple abstraction-level language for service-oriented applications in ubiquitous environments," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 16, no. 4, pp. 1–30, 2009.

[21] M. Brambilla and P. Fraternali, *Interaction flow modeling language: Model-driven UI engineering of web and mobile apps with IFML*. Morgan Kaufmann, 2014.

[22] M. Hamdani, W. H. Butt, M. W. Anwar, and F. Azam, "A systematic literature review on interaction flow modeling language (ifml)," in *Proceedings of the 2018 2nd International Conference on Management Engineering, Software Engineering and Service Sciences*, 2018, pp. 134–138.

[23] A. Huang, M. Pan, T. Zhang, and X. Li, "Static extraction of ifml models for android apps," in *Proceedings of the 21st ACM/IEEE International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings*, 2018, pp. 53–54.

[24] N. Yousaf, F. Azam, W. H. Butt, M. W. Anwar, and M. Rashid, "Automated model-based test case generation for web user interfaces (wui) from interaction flow modeling language (ifml) models," *IEEE Access*, vol. 7, pp. 67 331–67 354, 2019.

[25] R. K. Cao and X. Liu, "Ifml-based web application modeling," *Procedia Computer Science*, vol. 166, pp. 129–133, 2020.

[26] N. Kharmoum, S. Ziti, Y. Rhazali, and F. Omary, "An automatic transformation method from the e3value model to ifml model: An mda approach," *Journal of Computer Science*, vol. 15, no. 6, pp. 800–813, 2019.

[27] A. Moldovan, V. Nicula, I. Pasca, M. Popa, J. K. Namburu, A. Oros, and P. Brie, "Openuidl, a user interface description language for runtime omni-channel user interfaces," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. EICS, pp. 1–52, 2020.

[28] (2015) What is apache freemarker™? [Online]. Available: https://jonas-moennig.de/how-to-cite-a-website-with-bibtex/

[29] (2008) Template designer documentation¶. [Online]. Available: http://jinja.octoprint.org/templates.html

[30] mustache. [Online]. Available: https://mustache.github.io/

[31] xslt cover page. [Online]. Available: https://www.w3.org/TR/xslt/

Acceptance of YouTube for Islamic Information Acquisition: A Multi-group Analysis of Students'Academic Discipline

M. S Ishak[1]
Institute of Ethnic Studies
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor, Malaysia

A Sarkowi[2]
Institut Pendidikan Guru Kampus Darulaman
06000 Bandar Darulaman
Jitra, Kedah, Malaysia

M. F Mustaffa[3]
Lembaga Pengurusan Sekolah Zakat Kedah
Bangunan Kolej Sultan Abdul Halim
Alor Setar, Kedah, Malaysia

R Mustapha[4]
Academy of Contemporary Islamic Studies
Universiti Teknologi MARA Pahang
Raub Campus, Pahang, Malaysia

***Abstract*—YouTube has been recognized as an important information source for the millennial generation. This paper aims to identify the factors affecting Malaysian higher education students' acceptance of YouTube for Islamic information acquisition and to investigate if any notable distinction that exists between the students' path coefficients in Islamic academic discipline and the other disciplines. Employing the Unified Theory of Acceptance and Use of Technology model (UTAUT) as its theoretical foundation, data were collected by distributing a self-administered survey to 795 students actively using YouTube for information seeking. Partial least squares structural equation modelling (PLS-SEM) and multi-group analysis (MGA) in SmartPLS 3.2.7 software were used to analyze the data. Three constructs of the UTAUT model, performance expectancy, effort expectancy and social influences, were found to significantly and positively influence behavioural intention to use YouTube for Islamic information acquisition in both groups of students. Facilitating conditions demonstrates significantly negative relationship with YouTube acceptance for students in other academic disciplines than for Islamic academic discipline. Additionally, the MGA analysis' findings suggest that determinants' factor coefficients of YouTube acceptance for Islamic information acquisition are not significantly different between students in Islamic academic disciplines and the other disciplines. This study validates the UTAUT model to understand the determinant of social media application usage in a new study context.**

Keywords—Unified Theory of Acceptance and Use of Technology (UTAUT); YouTube; information acquisition; student knowledge; Partial Least Square

I. INTRODUCTION

YouTube resembles a social media platform that specializes in online videos. It is available in over 100 countries, with 400 hours of video added per minute in 80 languages [1]. With 2.3 billion users worldwide, the YouTube site has increased in popularity [2]. However, with no requirements or defined certification criteria, creators worldwide can submit an infinite number of videos to YouTube, meaning that video accuracy and quality might differ widely [3, 4]. This is because the content quality and content validity would depend on the knowledge and experience of video creators [5]. However, despite these facts, YouTube has been recognized as the first and essential source for millennial generations when they want to acquire new information and express their opinion [6-8]. Remarkably, the most popular search-related term on YouTube is "How to" [9].

There is an expanding literature body that has investigated YouTube as an information source predominantly pertaining to health-related information [10-13]. The findings of these studies, together with others, have highlighted the quality and reliability of shared health information. For example, Kocyigit et al. [13] and Ng et al. [11] found that most YouTube videos were high quality and provided useful information. However, several studies indicated that YouTube is not an appropriate source of information [10, 12]. Nevertheless, despite the fact of contrary evidence regarding specialized health information, the number of users obtaining health information from internet-based sources is rapidly increasing [13]. One explanation is that users have experienced enjoyable and valuable times when YouTube is used as a source of health information [14]. Nevertheless, studies on the usage of YouTube for Islamic information are scarcely found in the literature.

Traditionally, Islamic knowledge learning is usually performed through face-to-face interactions known as talaqqi and takes place in mosques, madrasahs, or other specific places [15, 16]. However, researchers have long been concerned about digital religious learning [17], which has motivated the present study. In general, as observed from prior studies, several studies highlighted the motivations of online Islamic information. For instance, Ishak [18] found that social media are uncomplicated and beneficial for Malaysian Muslims to obtain Islamic information to increase their religious beliefs. This is consistent with a recent study of Islamic information-seeking behavior in Saudi Arabia [19] and Indonesia [20].

However, most studies were conducted on general Internet or social media applications, and a specific study for YouTube has not been attempted.

The literature shows that UTAUT researching students' YouTube usage is being validated to describe and anticipate user behaviors behavioral and intention with regard to technology acceptance [21]–[26]. In contrast, Venkatesh et al. [28] proposed that it is crucial to evaluate UTAUT in a variety of technological contexts, settings, and cultural contexts since factors that affect technology adoption may differ depending on the technological context, cultural context, and user population. In this context, it was projected that the variables affecting students' use of YouTube for information acquisition could vary from contexts for using information systems generally, and that the UTAUT would need to be applied in diverse situations [26].

The objective of this study is twofold. First, to investigate the determinants of YouTube acceptance for Islamic information acquisition among Muslim university students in the margin of UTAUT. In order to fully grasp students' acceptance of YouTube, the four basic UTAUT constructs (performance expectancy, effort expectancy, social influence, and facilitating conditions) were utilized. Second, despite the numerous studies on YouTube adoption within various academic disciplines in Higher Education, such as medical sciences [29, 30], language literacy [31, 32], and business management [33, 34], the objective of our work is to fill the gap of the effect of academic disciplines on YouTube Islamic information acquisition.

This paper’s remaining section is like the following. As Section II describes the hypotheses development and conceptual framework, Section III explains the methodology for this study. Section IV discusses the results of the measurement model, structural model and multi-group analysis. Furthermore, Section V presents the study’s findings, and Section VI provides the paper’s conclusion.

II. Literature Review

A. Conceptual Framework

The Unified Theory of Acceptance and Use of Technology (UTAUT) model was chosen as the theoretical guideline for this research because it explains the elements that influence technology acceptance and usage [27]. Furthermore, several prior YouTube studies' usage in educational settings has empirically validated UTAUT for validity and reliability [23], [35–37]. Furthermore, UTAUT is a robust approach because it incorporates eight various models [27], encompassing the Theory of Reasoned Action [38, 39], Theory of Planned Behavior [40], Technology Acceptance Model [41], motivational model [42], the personal computer utilization model [43], an integrated model of planned behavior and technology acceptance [44], Social Cognitive Theory [45] as well as Innovation Diffusion Theory [46]. Moreover, facilitating conditions, social influence, effort expectancy, and performance expectancy are the four key variables that impact actual use and behavioral intention [27]. Therefore, three independent variables (social influence, effort expectancy, and performance expectancy) are suggested as behavioral intention direct determinants by UTAUT, whereas behavioral intention and facilitating condition determine actual usage. Hence, we suggest the conceptual framework provided in Fig. 1 for this research.

B. Hypotheses Development

Performance expectancy denotes a person's belief that technology will improve their performance in specific tasks [27]. Performance expectancy will be explored in connection to information acquisition activities in this research. The performance expectancy factor has been proven to possess a major effect on the behavioral intention of social media usage in education [23][47-50]. Thus, the research theorizes that the higher students perceive YouTube as an advantageous source of Islamic information acquisition, the higher their behavioral intention to use the application.

H1. Performance expectancy is a significant predictor of the behavioral intention to use YouTube for Islamic information acquisition.

The level of easiness connected with the technology used is referred to as effort expectancy [27]. Effort expectancy signifies a person's expectation of obtaining Islamic information from YouTube without much effort, as YouTube is a social media application. Researchers have found that the easiness of social media in educational settings positively influences the behavioral intentions to utilize the applications [47-49]. Hence, this research hypothesizes that the students' intention to utilize a YouTube video for Islamic information acquisition will increase with an advance in the searching easiness and handling of the YouTube functions.

H2. Effort expectancy is a significant predictor of the behavioral intention to use YouTube for Islamic information acquisition.

Social influence refers to a person's understanding of how others are relevant (e.g., peers, friends) and considers them utilizing technology [27]. Several research on social media acceptance has found that the influence of others is a strong predictor of social media behavioral intention for learning purposes [23][47-50].

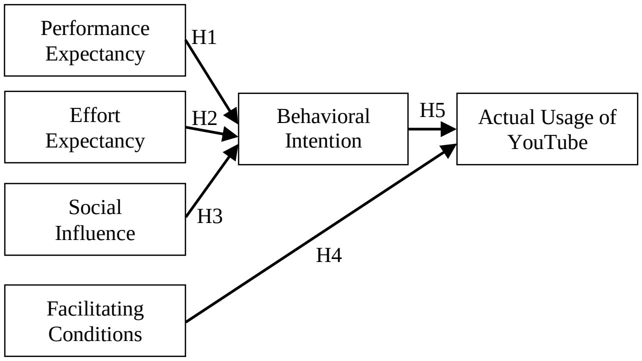


Fig. 1. The Research Framework and Hypotheses.

Therefore, individuals will be motivated to accept social media in educational settings if social influencers also utilize it. Hence, this work hypothesizes that the students' intentions to use YouTube for Islamic information acquisition increase when the social influence level increases (for instance, imposed by significant people, family, or peers).

H3. Social influence is a significant predictor of the behavioral intention to use YouTube for Islamic information acquisition.

The extent to which a person senses that technological infrastructure and an organizational sustains to ease the system's use is termed facilitating conditions [27]. Facilitating conditions were expressed in the context of this research as students' views of whether they have access to the tools and assistance they need to utilize YouTube for Islamic information acquisition. Several social media researchers have found that the facilitating conditions are a reliable predictor of intention to utilize social media apps [47][50]. Therefore, this research hypothesizes that when the ease of facilitating conditions improves, the student's use of YouTube videos for Islamic knowledge will develop.

H4: Facilitating conditions are a significant predictor of the behavioral intention to use YouTube for Islamic information acquisition.

Behavioral intention signifies the motivational factors that affect the probability of performing certain behavior [40]. The behavioral intention has been identified as a mediating element in users' use and acceptance of technology [27]. In the framework of this research, behavioral intention assesses students' inclinations, as well as intentions of using YouTube to get Islamic information. According to the UTAUT, behavioral intention is a powerful indicator of actual technology utilization. The behavioral intention has also been found to be a key predictor of social media application usage in several research on social media [23][47, 48][50]. Therefore, this study hypothesizes that behavioral intention is pivotal to YouTube's actual use of Islamic information.

H5: Behavioral intention is a significant predictor of students' actual usage of YouTube for Islamic information acquisition.

III. Methods

A. Samples

Referring to the UTAUT by Venkatesh et al. [27], this research adopts a quantitative method to assess all the variables in the study hypothesis model. The participants in the research are the younger generation of Muslims in Malaysia among undergraduate students at public universities aged between 20 to 30 years. The rationale for selecting university students is because they are active users of YouTube and utilize the platform for information seeking.

As per the Ministry of Higher Education Malaysia's report, there were 584,576 public university students in 2020 [51]. The sample size was selected according to the method prescribed for Partial Least Squares-Structural Equation Modelling (PLS-SEM) analysis. Based on the minimum sample size requirement suggested by Hair et al. [52], we would need a minimum 619 sample size to render the corresponding effect significant at 5% when the minimum path coefficient is expected to be signed between 0.11 and 0.20. This study targeted 800 students randomly selected from five areas in Malaysia at six public universities. Data were collected through online surveys assisted by the program coordinator in each selected university. Initially, 800 surveys were collected; 795 students were used for the analysis after removing the outliers.

B. Measures

The questionnaire consists of demographics, YouTube usage patterns, and five variables taken from UTAUT and administered in the Malay language. The measurement instrument is presented in Appendix A. The interval scale was used to six variables measure: performance expectancy (Cronbach’s Alpha=0.886), effort expectancy (Cronbach’s Alpha=0.935), social influence (Cronbach’s Alpha=0.917), facilitating conditions (Cronbach’s Alpha=0.939), behavioral intentions (Cronbach’s Alpha=0.926), as well as actual usage (Cronbach’s Alpha=0.909). In the pre-test research, they all had high construct reliability.

C. Data Analysis

The partial least squares structural equation modelling (PLS-SEM) was used for this investigation due to its ideal for theory prediction and exploration [53]. SmartPLS version 3.2.7 [54] was used. The following actions were taken during data analysis: to see if any demographic variables differed between Islamic and other academic disciplines, a chi-squared test was utilized alongside IBM SPSS 26.0. The model was subsequently verified using a two-stage analytical PLS-SEM procedure [55]. PLS-SEM has the additional benefit of evaluating both the measurement and structural models, which is additionally best suited to performing multi-group analysis [54][57].

The first stage involves testing the measurement model to assess the internal consistency reliability, convergent validity and discriminant validity of the constructs used in this study. Composite reliability of more than 0.70 [55] was used to analyze the reliability. Compared to Cronbach's Alpha, composite reliability is better suited for PLS-SEM [53]. In addition, the average variance extracted, and factor loading was utilized to evaluate convergent validity. Meanwhile, the Heterotrait-Monotrait ratio (HTMT) and the Fornell-Larcker criterion were used to determine discriminant validity [56]. Finally, an invariance test was also carried out to identify if the construct measurements were equally comprehended all across two groups of different academic disciplines before performing a multi-group analysis (MGA).

The second stage involves examining the structural model to test the hypotheses. The structural model defined the casual relationships between the model's constructs (the coefficient of determination, R^2 value and path coefficients). The R^2 and the path coefficients (significance and beta) together demonstrate how well the data are consistent with the proposed model [55][57]. The bootstrapping method involving a resampling of 5000 was employed to calculate the path coefficient's significance. According to the literature, this study additionally evaluated the path model's predictive relevance (Q^2 value), which includes blindfolding procedure [55][57]. Additionally, effect sizes (f^2) were evaluated to identify if an exogenous latent construct has a weak, moderate, or substantial influence on an endogenous latent construct [52]. Lastly, the multi-group analysis is carried out to establish the differences in path coefficients between two groups.

IV. Results

A. Sample Profiles

The chi-squared test discovered substantial differences (p<0.05) between students in Islamic academic discipline and other academic disciplines in gender, university, year of study, and age (Table I). Questions with multiple-choice answers were used to examine YouTube search behavior is shown in Table II.

TABLE I. Sample Profiles

Variables	***Profiles***	**Full Dataset n (%)**	**Islamic Academic Discipline Dataset n (%)**	**Other Academic Disciplines Dataset n (%)**	***Chi-Square***	***p Value***
Gender	Male	317 (39.9)	97 (30.6)	220 (69.4)	18.357	0.000
	Female	478 (60.1)	202 (42.3)	276 (57.7)		
University	USIM	200 (25.1)	99 (49.5)	101 (50.5)	300.509	0.000
	UMK	100 (12.6)	0 (0.0)	100 (100.0)		
	UMS	100 (12.6)	0 (0.0)	100 (100.0)		
	UKM	100 (12.6)	100 (100.0)	0 (0.0)		
	UNIMAP	100 (12.6)	0 (0.00)	100 (100.0)		
	UIA	195 (24.5)	100 (51.3)	95 (48.7)		
Study Year	Year 1	167 (21.0)	70 (41.9)	97 (58.1)	21.813	0.000
	Year 2	204 (25.7)	53 (26.0)	151 (74.0)		
	Year 3	291 (36.6)	107 (36.8)	184 (63.2)		
	Year 4	133 (16.7)	69 (51.9)	64 (48.1)		
Age	18 to 21	517 (53.0)	124 (24.0)	393 (76.0)	38.935	0.001
	22 to 24	402 (41.2)	161 (40.0)	241 (60.0)		
	25 to 27	48 (4.92)	13 (27.0)	35 (73.0)		
	28 to 30	8 (0.82)	1 (12.5)	7 (87.5)		

Almost all students actively used YouTube to get religious information based on self-initiative (94.0%). Searching keywords are based on topics (54.6%), current issues (51.2%), religious figures (45.9%), and religious law (33.0%). Students actively search YouTube on campus (77.6%), at home (58.3%), and others (10.8%) using smartphones (79.2%), notebooks (46.4%), and tablet (12.3%).

B. Measurement Invariance Test

A measurement invariance test assesses if item measurements vary between groups [55].

TABLE II. YouTube Searching Behavior

Variables	**Full Dataset n (%)**	**Islamic Academic Discipline Dataset n (%)**	**Other Academic Disciplines Dataset n (%)**
Using YouTube for religious information on own initiatives:	748 (94.0)	290 (38.8)	458 (61.2)
Keywords used for searching religious information:			
Topics	434 (54.6)	175 (40.3)	259 (59.7)
Current Issues	407 (51.2)	163 (40.0)	245 (60.0)
Religious Figures	365 (45.9)	141 (38.6)	224 (61.4)
Religious Law	262 (33.0)	108 (41.2)	154 (58.8)
Access YouTube from:			
Campus Network	617 (77.6)	261 (42.3)	356 (57.7)
Home	464 (58.3)	163 (35.1)	301 (64.9)
Others	86 (10.8)	30 (34.9)	56 (65.1)
Device Accessing YouTube:			
Smartphone	630 (79.2)	247 (39.2)	383 (60.8)
Notebook	369 (46.4)	156 (42.3)	213 (57.7)
Tablet	98 (12.3)	32 (32.7)	66 (67.3)

When executing multi-group analysis, this is a key step [58]. The initial examination found that full measurement invariance was not possible. We discovered that one indicator of effort expectancy (EE2) has distinct implications across the groups and excluded this indication. All construct measures were invariant between Islamic academic discipline and other academic disciplines dataset once the initial model was purified; there was no substantial distinction between those two. The findings of the investigation revealed well-developed measurement invariance (Table III).

C. Measurement Model

Assessment of the measurement model was done through construct reliability as well as validity.

For construct reliability, this study tested the individual composite reliability (CR) values to evaluate the reliability of each of the measurement model's key variables [55]. The results indicate that all the composite reliability (CR) values range from 0.799 to 0.894 in the full data set, 0.773 to 0.94 in the Islamic discipline group, and 0.827 to 0.938 in the other disciplines group, respectively. These values were higher than 0.7 [39], which, as demonstrated in Table IV, sufficiently demonstrates that construct reliability is met.

TABLE III. MEASUREMENT INVARIANCE TEST

Construct	**Indicator**	**Outer Loadings Differences**	***p Value***
Performance Expectancy (PE)	PE1	0.023	0.372
	PE2	0.023	0.225
	PE3	0.245	0.669
	PE4	0.285	0.636
	PE5	0.001	0.969
Effort Expectancy (EE)	EE1	0.128	0.619
	EE3	0.049	0.847
	EE4	0.216	0.917
	EE5	0.209	0.178
Social Influence (SI)	SI1	0.185	0.537
	SI2	0.189	0.677
	SI3	0.200	0.287
	SI4	0.281	0.148
Facilitating Conditions (FC)	FC1	0.056	0.403
	FC2	0.101	0.875
	FC3	0.021	0.735
	FC4	0.111	0.813
	FC5	0.042	0.739
Behavioral Intention (BI)	BI1	0.006	0.685
	BI2	0.025	0.221
	BI3	0.005	0.692
	BI4	0.024	0.538
Actual Usage (AU)	AU	0.000	0.671

Indicator reliability was examined using factor loading. High loadings on a construct show that the related indicators appear to have similarities which what the construct captures [55]. Factor loadings of 0.50 were regarded as highly significant [55]. As indicated in Table IV, all items' loadings were higher than the recommended value of 0.5, with the exception of item FC2 in three different data sets and SI4 in the Islamic academic discipline group. These items were retained because they not affect internal consistency and reliability [52]. However, items EE1 and FC4 were deleted to improve the composite reliability.

This study employed the average variance extracted (AVE) to examine convergent validity (the degree to which a measure correlates favourably with different measures of the same concept), and it found that all AVE values exceeded the recommended value of 0.50 [55]. As a result, the convergent validity for all constructs in the three datasets has been well met. Thus, sufficient convergent validity is exhibited, as shown in Table IV. The full dataset indicates values of AVE from 0.551 to 0.793, Islamic academic discipline data pointed the values ranging from 0.540 to 0.797, and other academic disciplines data indicated the values of AVE from 0.556 to 0.791.

Two metrics, Fornell-Larcker and the Heterotrait-Monotrait (HTMT) ratio, were used to evaluate the discriminant validity of the measurement model (the extent to which items differ between constructs or measure various concepts). The results of discriminant validity using the Fornell-Larcker criterion for the full dataset are presented in Table V, in which the square root of the AVEs on the diagonals, as denoted by the bolded values, are higher than the correlations between constructs (column values and corresponding row). This suggests that the constructs have an excellent discriminant validity because they are substantially associated to their respective indicators when compared to other model constructs [59]. Moreover, the correlation between exogenous constructs is lower than 0.85 [59]. Additionally, the constructs' Fornell-Larcker criterion for the Islamic academic discipline dataset (Table VI) and other academic disciplines dataset (Table VII) also suggested good discriminant validity.

TABLE IV. CONVERGENT VALIDITY AND RELIABILITY

Construct	**Indicator**	**Factor Loadings**			**Composite Reliability (CR)**			**Average Variance Extracted (AVE)**		
		Full Dataset	***Islamic AcademicDiscipline Dataset***	***Other Academic DisciplinesDataset***	***Full Dataset***	***Islamic AcademicDiscipline Dataset***	***Other Academic DisciplinesDataset***	***Full Dataset***	***Islamic Academic Discipline Dataset***	***Other Academic DisciplinesDataset***
PE	PE1	0.854	0.869	0.846	0.894	0.940	0.873	0.630	0.758	0.586
	PE2	0.889	0.906	0.883						
	PE3	0.744	0.909	0.664						
	PE4	0.626	0.834	0.549						
	PE5	0.828	0.829	0.831						
EE	EE3	0.705	0.549	0.884	0.799	0.773	0.861	0.571	0.540	0.677
	EE4	0.731	0.890	0.648						
	EE5	0.825	0.726	0.911						
SI	SI1	0.792	0.927	0.742	0.828	0.855	0.862	0.551	0.612	0.611
	SI2	0.801	0.719	0.907						
	SI3	0.798	0.933	0.733						
	SI4	0.546	0.449	0.730						

FC	FC1	0.854	0.882	0.826	0.882	0.828	0.827	0.563	0.564	0.556
	FC2	0.440	0.377	0.476						
	FC3	0.878	0.858	0.881						
	FC5	0.747	0.776	0.733						
BI	BI1	0.897	0.901	0.895	0.830	0.940	0.938	0.793	0.797	0.791
	BI2	0.905	0.920	0.895						
	BI3	0.928	0.931	0.926						
	BI4	0.829	0.815	0.839						
AU	B2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE V. DISCRIMINANT VALIDITY OF THE FULL DATASET

Fornell-Larcker Criterion							**Heterotrait-Monotrait Ratio (HTMT)**						
	AU	***BI***	***EE***	***FC***	***PE***	***SI***		***AU***	***BI***	***EE***	***FC***	***PE***	***SI***
AU	**1.000**						AU						
BI	-0.098	**0.890**					BI	0.101					
EE	-0.062	0.576	**0.756**				EE	0.076	0.749				
FC	-0.117	0.550	0.411	**0.750**			FC	0.133	0.656	0.593			
PE	-0.081	0.655	0.638	0.448	**0.794**		PE	0.090	0.732	0.861	0.547		
SI	-0.054	0.561	0.577	0.329	0.522	**0.742**	SI	0.061	0.683	0.849	0.452	0.660	

Notes: PE: performance expectancy; SI: social influence; EE: effort expectancy; FC: facilitating condition; AU: actual usage; BI: behavioral intention

TABLE VI. DISCRIMINANT VALIDITY OF ISLAMIC ACADEMIC DISCIPLINE DATASET

Fornell-Larcker Criterion							**Heterotrait-Monotrait Ratio (HTMT)**						
	AU	***BI***	***EE***	***FC***	***PE***	***SI***		***AU***	***BI***	***EE***	***FC***	***PE***	***SI***
AU	**1.000**						AU						
BI	-0.138	**0.893**					BI	0.138					
EE	-0.062	0.629	**0.735**				EE	0.102	0.801				
FC	-0.117	0.609	0.535	**0.751**			FC	0.136	0.712	0.759			
PE	-0.081	0.686	0.647	0.531	**0.870**		PE	0.103	0.744	0.837	0.627		
0SI	-0.054	0.643	0.620	0.410	0.579	**0.782**	SI	0.061	0.732	0.894	0.514	0.655	

Notes: PE: performance expectancy; EE: effort expectancy; SI: social influence; FC: facilitating condition; BI: behavioral intention; AU: actual usage

TABLE VII. DISCRIMINANT VALIDITY OF OTHER ACADEMIC DISCIPLINES DATASET

Fornell-Larcker Criterion							**Heterotrait-Monotrait Ratio (HTMT)**						
	AU	***BI***	***EE***	***FC***	***PE***	***SI***		***AU***	***BI***	***EE***	***FC***	***PE***	***SI***
AU	**1.000**						AU						
BI	-0.074	**0.893**					BI	0.077					
EE	-0.090	0.629	**0.735**				EE	0.093	0.801				
FC	-0.111	0.609	0.535	**0.751**			FC	0.127	0.712	0.759			
PE	-0.079	0.686	0.647	0.531	**0.870**		PE	0.089	0.744	0.837	0.627		
SI	-0.076	0.643	0.620	0.410	0.579	**0.782**	SI	0.083	0.732	0.894	0.514	0.655	

Notes: PE: performance expectancy; EE: effort expectancy; SI: social influence; FC: facilitating condition; BI: behavioral intention; AU: actual usage.

However, Henseler et al. [56] mentioned that the lack of discriminant validity in most study scenarios is not accurately revealed by the Fornell-Larcker criterion. Therefore, a new approach has been introduced to solve this sensitivity problem, which uses the HTMT ratio of correlations based on the multitrait-multimethod matrix. The results for discriminant validity using the HTMT ratio are presented in Table V (full dataset), Table VI (Islamic academic discipline dataset) and Table VII (other academic disciplines dataset).The conservative approach’s HTMT threshold is <0.85, whereas for the liberal approach’s threshold is <0.90 [38]. The liberal approach was used in the current study to evaluate discriminant validity. The findings demonstrate a lack of discriminant validity since all construct values in the three datasets are less than 0.90. In conclusion, based on the tests of both validity and reliability, the measurement model in three datasets has been validated successfully.

D. Structural Model

The structural model defined the causal relationships between the constructs in the model (the coefficient of determination, R^2 value and path coefficients). Both the path coefficients (significance and β) and the R^2 display excellent ways that the data is supporting the hypothesized model [55][57]. Table VIII presents the path coefficients for the entire sample. After analysis, it was discovered that behavioural intention (BI) was positively and significantly associated to effort expectancy (EE), performance expectancy (PE), and social influence (SI) for all datasets. Table VIII also displays the results in detail for each academic discipline. Most notably, both the academic disciplines datasets and the entire datasets offer empirical support for H1, H2, and H3, respectively.

With regard to the relationship between facilitating condition (FC) and actual usage (AU), the full and other academic disciplines datasets were discovered to possess a negative significant outcome. Thus, H4 is strongly supported, except for the dataset for Islamic academic discipline. The R^2 values in Table IX for this relationship were above the desired 0.5 threshold and considered moderate [55], accounting for 51% of behavioural intention (BI) variance. However, there is no empirical support for the impact of behavioural intention (BI) on actual usage (AU) for all datasets.

This study also evaluated the path model's predictive relevance (Q^2 value) via the blindfolding procedure. Every data point of the indicators in the reflective measurement model of endogenous constructs is systematically deleted and predicted during the blindfolding procedure, which is a resampling technique [55]. The original values and the forecasted values are compared using this procedure. The path model has good predictive accuracy if the prediction is close to the original values (for example, the prediction error is minimal). If the value of Q^2 is greater than 0, then the predictive relevance of the proposed model exists for a certain endogenous construct [55], [59]. The findings for all datasets reveal that the Q^2 values are all greater than 0, verifying the prediction model as shown in Table IX.

Finally, this study assessed the f^2 effect sizes. It identifies if an exogenous latent construct possess a weak, moderate, or substantial impact on an endogenous latent construct [51]. Hair et al. [55] suggest testing the R^2 value's changes. On the other hand, Cohen [61] recommend a magnitude of f^2 at 0.35 (large effects), 0.15 (medium effects) and 0.02 (small effects) as a guideline measure. Following the rules of thumb, the result of f^2 as Table IX shows the relationships with large effects on behavioural intention (BI) of the total dataset but small effect sizes on actual usage (AU).

TABLE VIII. STRUCTURAL ESTIMATES

Hypotheses	**Full Dataset**			**Islamic Academic Discipline Dataset**			**Other Academic Disciplines Dataset**		
	Path Coefficients (β)	***p Value***	***Results***	***Path Coefficients (β)***	***p Value***	***Results***	***Path Coefficients (β)***	***p Value***	***Results***
H1: PE → BI	0.402	0.000	S	0.373	0.000	S	0.305	0.000	S
H2: EE → BI	0.194	0.000	S	0.208	0.001	S	0.291	0.000	S
H3: SI → BI	0.239	0.000	S	0.298	0.000	S	0.225	0.000	S
H4: FC → AU	-0.090	0.019	S	-0.072	0.293	NS	-0.100	0.027	S
H5: BI → AU	-0.048	0.250	NS	-0.095	0.169	NS	-0.023	0.658	NS

Notes: S: supported; NS: not supported; PE: performance expectancy; EE: effort expectancy; SI: social influence; FC: facilitating condition; BI: behavioral intention; AU: actual usage

TABLE IX. RESULTS OF R^2, Q^2, AND F^2 ANALYSIS

Construct	**Full Dataset**				**Islamic Academic Discipline Dataset**				**Other Academic Disciplines Dataset**			
	R^2	***Q^2***	***f^2***		***R^2***	***Q^2***	***f^2***		***R^2***	***Q^2***	***f^2***	
			AU	***BI***			***AU***	***BI***			***AU***	***BI***
EE	-	-		0.038	-	-		0.048	-	-		0.071
FC	-	-	0.006		-	-	0.003		-	-	0.008	
PE	-	-		0.177	-	-		0.167	-	-		0.084
SI	-	-		0.075	-	-		0.120	-	-		0.065
BI	0.513	0.401	0.002		0.582	0.456	0.006		0.512	0.396	0.000	
AU	0.015	0.011			0.022	0.008			0.013	0.006		

Notes: PE: performance expectancy; SI: social influence; EE: effort expectancy; FC: facilitating condition; BI: behavioral intention; AU: actual usage

E. Multi-group Analysis (MGA)

Using MGA, the analysis concludes by determining if the differences in path coefficients between the Islamic academic discipline dataset and other academic disciplines dataset are significant. The results in Table X indicate no substantial differences between the two groups were observed in any pathways. This indicates that there is no significant difference between the two different groups of academic disciplines in terms of the effect of performance expectancy (PE), effort expectancy (EE), and social influence (SI) on behavioural intention (BI) to use YouTube for Islamic information acquisition. This study also indicates that there is no significant difference between two different groups of academic disciplines in terms of the effect of facilitating condition (FC) and behavioural intention (BI) on the actual usage (AU) of YouTube for Islamic information acquisition.

V. Discussion

This study used the UTAUT model as a guiding concept to analyze the factors that influence students' behavioral intentions to utilize YouTube for Islamic information acquisition.

TABLE X. Path Differences by Academic Disciplines

Relationships	**Path Coefficients Differences**	***t Value***	***p Value***
PE → BI	0.100	1.199	0.231
EE → BI	0.117	1.316	0.189
SI → BI	0.077	1.030	0.303
FC → AU	0.027	0.336	0.737
BI → AU	0.071	0.852	0.394

Notes: PE: performance expectancy; SI: social influence; EE: effort expectancy; FC: facilitating condition; BI: behavioral intention; AU: actual usage.

It also looked at how academic disciplines affected these relationships. The PLS-SEM approach's empirical outcomes showed that the relevance and relative significance of influencing elements in the UTAUT on behavioral intention and usage is not distinguished between students in Islamic and other academic disciplines.

Significant relationships exist between performance expectations and students' behavioral intentions to use YouTube for Islamic information acquisition, supporting H1 as anticipated. The path coefficients did not vary significantly between the Islamic academic discipline and other academic discipline groups of students. The easy access to YouTube among students might explain that video on Islamic material is available anywhere, anytime, on various devices. Hence, this result aligns with earlier studies that used UTAUT in YouTube research contexts for information acquisition [23].

Parallel to the hypothesized relationship, effort expectancy affects students' use of YouTube for Islamic Information acquisition for both groups in different academic disciplines. Thus, H2 is upheld. This result has also been a decisive factor in the intention to use social media for learning in previous studies [47-49]. For this study, it may be assumed that the students' use of YouTube for Islamic learning does not necessitate any intellectual, physical, emotional, or psychological effort on their part. In other words, YouTube apps are simple to use for information acquisition.

Regarding social influence on behavioural intention to use YouTube for Islamic information, this research's findings portrayed that students do emphasize this, and hence H3 is supported. This research's results are consistent with previous studies in UTAUT, which explain that social influence factors are closely related to a person's behavioral intentions to use YouTube [23][47-50]. Furthermore, the effect of social influences on the YouTube use intention was greater for students in Islamic discipline than for other academic disciplines. This interesting result might come from an increased number of Muslim preachers utilizing YouTube to spread their message [62, 63]. This study, even so, found that the difference in path coefficients between the two groups was insignificant.

The existence of facilitating conditions positively affects students' actual usage of YouTube for Islamic Information overall and other academic disciplines students. Hence, H4 was sustained. However, it was not a determinant for Islamic academic disciplines students. This predictor's path coefficients on YouTube usage appeared to be higher among other academic disciplines students. Nonetheless, in this study, there were no significant differences in the path coefficients of the two groups. This finding is both enlightening and controversial. The explanation for this might be due to the enabling condition's limited effect size on other academic disciplines of student usage. Furthermore, there were no variations in the effect size of the facilitating conditions among the two groups. The result is along the lines of earlier literature [23] that found that facilitating condition was not a predictor of students' willingness to use YouTube application for learning.

Unlikely, the behavioral intention was not significantly related to the YouTube usage for either Islamic or other academic discipline students; there was no substantial difference between the two groups. This intriguing outcome might be since using YouTube to obtain Islamic material is a voluntary use of technology. However, several research that deemed social media use for learning to be mandatory disputed these findings [23][47].

This research presents a statistically verified UTAUT model to explain the different usage between students of Islamic and other academic disciplines. In the context of disseminating Islamic information on online platforms, this study corroborates the UTAUT's ability to predict students' behavioral intentions regardless of their academic background. The students of nowadays are known as ‘digital natives,’ who are accustomed to incorporating technology into every aspect of their lives, including religious activities. They exploited YouTube as a search engine to discover information in formal and informal learning [8]. However, previous research discovered obstacles to misinformation and disinformation of Islamic information on social media [20]. Given these results, Muslim preachers should utilize this platform to disseminate Islamic information as the video's credibility depends on them. Furthermore, YouTube potentially encourages interactivity among users and preachers because it has tools that facilitate interaction.

Finally, this study has limitations that may suggest future research directions. First, possible latent variables that may enhance the prediction of students' behavioral intentions, such as ambiguity of information [64] and students' religiosity, should be considered in future research. Second, the study examined just Malaysian students; results obtained in other countries may differ. Lastly, this research depended on respondents' self-reported measures of actual usage, which might be skewed by common method bias. Assessing students' usage in a longitudinal study and controlling for real YouTube usage would be a more precise estimate.

VI. Conclusion

This research's primary goal was to observe the factors influencing students' behavior when it comes to YouTube usage for Islamic information acquisition established on the UTAUT research framework. This research also analyzes the differences between Islamic academic disciplines and other academic disciplines groups of students in the usage of YouTube for Islamic information acquisition. The findings imply that for both groups, students' behavioral intentions to YouTube for these specific purposes are predicted by social influences, effort expectancy, and performance expectancy. The results also found that that facilitating conditions were not a significant predictor of Islamic academic disciplines students' actual YouTube usage. Nonetheless, for both groups, behavioral intention was not a substantial predictor of actual YouTube usage for Islamic information acquisition. Overall, results of the multi-group analysis of PLS-SEM showed that there were no substantial differences were observed in any path coefficient between the two groups. The research's theoretical contribution is to verify the UTAUT model in the usage of YouTube among students in higher education for specific information acquisition.

Acknowledgment

Acknowledgement to the Ministry of Higher Education Malaysia, Research Acculturation Grant (RAGS) S/O Code: 12854 on the usage of YouTube among the young Muslim generation in Malaysia: The application of the Unified Theory of Acceptance and Use of Technology (UTAUT).

References

[1] YouTube, “YouTube by the number,” YouTube official blog. https://blog.youtube/press/.

[2] Statista, “Most popular social networks worldwide as of January 2022, ranked by number of monthly active users,” Oct. 2021. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (accessed Dec. 21, 2021).

[3] M. Farag, D. Bolton, and N. Lawrentschuk, “Use of youtube as a resource for surgical education-clarity or confusion,” Eur. Urol. Focus, vol. 6, no. 3, pp. 445–449, May 2020, doi: 10.1016/j.euf.2019.09.017.

[4] T. Godskesen, S. F. Holm, A. T. Höglund, and S. Eriksson, “YouTube as a source of information on clinical trials for paediatric cancer,” Information, Commun. Soc., pp. 1–14, Sep. 2021, doi: 10.1080/1369118X.2021.1974515.

[5] A. K. Khilnani, R. Thaddanee, and G. Khilnani, “‘Learning ENT’by YouTube videos: Perceptions of third professional MBBS students,” Int. J. Otorhinolaryngol. Head Neck Surg., vol. 6, no. 6, p. 1120, Jun. 2020, doi: 10.18203/issn.2454-5929.ijohns20202211.

[6] Y. Taskin, T. Hecking, H. U. Hoppe, V. Dimitrova, and A. Mitrovic, “Characterizing comment types and levels of engagement in video-based learning as a basis for adaptive nudging,” in Proceeding of the European Conference on Technology Enhanced Learning, 2019, pp. 362–376.

[7] D. Zimmermann et al., “Influencers on YouTube: A quantitative study on young people’s use and perception of videos about political and societal topics,” Curr. Psychol., vol. 2020, pp. 1–17, Nov. 2020, doi: https://doi.org/10.1007/s12144-020-01164-7.

[8] F. Pires, M. J. Masanet, and C. A. Scolari, “What are teens doing with YouTube? Practices, uses and metaphors of the most popular audio-visual platform,” Information, Commun. Soc., vol. 24, no. 9, pp. 1175–1191, Jul. 2021, doi: 10.1080/1369118X.2019.1672766.

[9] S. Utz and L. N. Wolfers, “How-to videos on YouTube: The role of the instructor,” Information, Commun. Soc., vol. 2020, pp. 1–16, May 2020, doi: 10.1080/1369118X.2020.1804984.

[10] F. Ayranci, S. K. Buyuk, and K. Kahveci, “Are YouTubeTM videos a reliable source of information about genioplasty?,” J. Stomatol. Oral Maxillofac. Surg., vol. 122, no. 1, pp. 39–42, Feb. 2021, doi: 10.1016/j.jormas.2020.04.009.

[11] C. H. Ng, G. R. S. Lim, and W. Fong, “Quality of English‐language videos on YouTube as a source of information on systemic lupus erythematosus,” Int. J. Rheum. Dis., vol. 23, no. 12, pp. 1636‐1644, Dec. 2020, doi: 10.1111/1756-185X.13852.

[12] C. Thomas, J. Westwood, and G. F. Butt, “Qualitative assessment of YouTube videos as a source of patient information for cochlear implant surgery,” J. Laryngol. Otol., vol. 135, no. 8, pp. 671–674, Aug. 2021, doi: 10.1017/S0022215121001390.

[13] B. F. Kocyigit, M. S. Akaltun, and A. R. Sahin, “YouTube as a source of information on COVID-19 and rheumatic disease link,” Clin. Rheumatol., vol. 39, no. 7, pp. 2049–2054, Jul. 2020, doi: 10.1007/s10067-020-05176-3.

[14] R. Al-Maroof et al., “The acceptance of social media video for knowledge acquisition, sharing and application: A comparative study among YouYube users and TikTok users’ for medical purposes,” Int. J. Data Netw. Sci., vol. 5, no. 3, p. 197, 2021, doi: 10.5267/j.ijdns.2021.6.013.

[15] N. A. Abdullah, Z. Temyati, and M. N. Mamat, “Technology related tahfiz Al-Quran learning in Malaysia: A systematic literature review,” Sains Insa., vol. 6, no. 3, pp. 47–52, Dec. 2021, doi: 10.33102/sainsinsani.vol6no3.361.

[16] F. R. Moeis, “Unraveling the myth of madrasah formal education quality in Indonesia: A labor quality approach,” Educ. Res. Policy Pract., pp. 1–24, Jun. 2021, doi: 10.1007/s10671-021-09298-6.

[17] D. Solahudin and M. Fakhruroji, “Internet and Islamic learning practices in Indonesia: Social media, religious populism, and religious authority,” Religions, vol. 11, no. 1, p. 19, Dec. 2019, doi: 10.3390/rel11010019.

[18] M. S. Ishak, “Pemodelan penerimagunaan maklumat berkaitan Islam di Internet: Pengaplikasian Model Penerimaan Teknologi (TAM),” J. Techno-Social, vol. 6, no. 2, pp. 49–61, Aug. 2014.

[19] A. Almobarraz, “Investigating the seeking behavior for religious information in social media,” in Research anthology on religious impacts on society, M. Khosrow-Pour, Ed. Hershey, PA: IGI Global, 2021, pp. 765–779.

[20] H. F. Amrullah, M. N. S. Ali, and M. F. Sukimi, “Information-seeking behavior of college students on religious tolerance through social media,” Islam. Int. J. Islam. Stud., vol. 41, no. 2, pp. 9–15, Jul. 2019, doi: 10.17576/islamiyyat-2019-4001-02.

[21] A. M. Al-Rahmi, A. Shamsuddin, and O. A. Alismaiel, “Unified theory of acceptance and use of technology (UTAUT) Theory: The Factors Affecting Students’ Academic Performance in Higher Education,” Psychol. Educ., vol. 57, no. 9, pp. 2839–2848, 2020.

[22] J. B. Awotunde, F. Roseline Oluwaseun Ogundokun, emi E. Ayo, G. J. Ajamu, and O. E. Ogundokun, “UTAUT model: Integrating social media for learning purposes among university students in Nigeria,” SN Soc. Sci., vol. 1, no. 9, pp. 1–27, Sep. 2021, doi: 10.1007/s43545-021-002.

[23] S. Bardakcı, “Exploring high school students’ educational use of YouTube,” Int. Rev. Res. Open Distrib. Learn., vol. 20, no. 2, pp. 260–278, Apr. 2019, doi: 10.19173/irrodl.v20i2.4074.

[24] M. Habes, M. Elareshi, and A. Ziani, “An empirical approach to understanding students’ academic performance: YouTube for learning

during the Covid-19 pandemic," Linguist. Antverp., vol. 2021, no. 3, pp. 1518–1534, Jul. 2021.

[25] I. Jung and Y. Lee, "YouTube acceptance by university educators and students: A cross-cultural perspective," Innov. Educ. Teach. Int., vol. 52, no. 3, pp. 243–253, May 2015, doi: 10.1080/14703297.2013.805986.

[26] P. Ramírez-Correa, A. Mariano-Melo, and J. Alfaro-Pérez, "Predicting and explaining the acceptance of social video platforms for learning: The case of Brazilian YouTube users," Sustainability, vol. 11, no. 24, p. 7115, 2019, doi: 10.3390/su11247115.

[27] V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology," MIS Q., vol. 2012, pp. 157–178, Mar. 2012, doi: 10.2307/30036540.

[28] S. Bardakcı. "Exploring high school students' educational use of YouTube," Int. Rev. of Res. in Opn. and Dist. Learn. vol. 20, no. 2, April 2019, doi: 10.19173/irrodl.v20i2.4074.

[29] S. M. Attardi, D. J. Gould, R. L. Pratt, and V. A. Roach, "YouTube‐based course orientation videos delivered prior to matriculation fail to alleviate medical student anxiety about anatomy," Anatomical Sciences Education, Research Report No: ASE-20-0433.R1, Jul. 2021. doi: 10.1002/ase.2107.

[30] O. Nomura, J. Irie, Y. Park, H. Nonogi, and H. Hanada, "Evaluating effectiveness of YouTube videos for teaching medical students CPR: Solution to optimizing clinician educator workload during the COVID-19 pandemic," Int. J. Environ. Res. Public Health, vol. 18, no. 13, p. 7113, Jul. 2021, doi: 10.3390/ijerph18137113.

[31] M. A. Sangermán Jiménez, P. Ponce, and E. Vázquez-Cano, "YouTube videos in the virtual flipped classroom model using brain signals and facial expressions," Futur. Internet, vol. 13, no. 9, p. 224, Sep. 2021, doi: 10.3390/fi13090224.

[32] S. Kim and H. C. Kim, "The benefits of YouTube in learning English as a second language: A qualitative investigation of Korean freshman students' experiences and perspectives in the US," Sustainability, vol. 13, no. 13, p. 7365, Jan. 2021, doi: 10.3390/su13137365.

[33] D. Schulz, A. van der Woud, and J. Westhof, "The best indycaster project: Analyzing and understanding meaningful YouTube content, dialogue and commitment as part of responsible management education," Int. J. Manag. Educ., vol. 18, no. 1, p. 100335, Mar. 2020, doi: 10.1108/JMD-03-2020-0087.

[34] F. Parra, A. Jacobs, and L. L. Trevino, "Shippy Express: Augmenting accounting education with Google Sheets," J. Account. Educ., vol. 56, p. 100740, Sep. 2021, doi: 10.1016/j.jaccedu.2021.100740.

[35] A. S. Al-Adwan, H. Yaseen, A. Alsoud, F. Abousweilem, and W. M. Al-Rahmi, "Novel extension of the UTAUT model to understand continued usage intention of learning management systems: the role of learning tradition," Educ. Inf. Technol., vol. 2021, pp. 1–27, Sep. 2021, doi: 10.1007/s10639-021-10758-y.

[36] M. M. M. Abbad, "Using the UTAUT model to understand students' usage of e-learning systems in developing countries," Educ. Inf. Technol., vol. 26, no. 6, pp. 7205–7224, Nov. 2021, doi: 10.1007/s10639-021-10573-5.

[37] T. A. Oteyola, T. A. Bada, and I. O. Akande, "Southwestern Nigerian University Undergraduates' acceptance of YouTube as a web-based instructional tool," Adv. Soc. Sci. Res. J., vol. 6, no. 8, pp. 45–57, Aug. 2019, doi: 10.14738/assrj.68.6866.

[38] I. Ajzen and M. Fishbein, Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[39] M. Fishbein and I. Ajzen, Belief, attitude, intentions and behavior: An introduction to theory and research. Boston, MA: Addison-Wesley, 1975.

[40] I. Ajzen, "The theory of planned behavior," Organ. Behav. Hum. Decis. Process., vol. 50, no. 2, pp. 179–211, Dec. 1991.

[41] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," MIS Q., vol. 13, no. 3, pp. 319–340, Sep. 1989, doi: 10.2307/249008.

[42] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "Extrinsic and intrinsic motivation to use computers in the workplace 1," J. Appl. Soc. Psychol., vol. 22, no. 14, pp. 1111–1132, Jul. 1992, doi: 10.1111/j.1559-1816.1992.tb00945.x.

[43] R. L. Thompson, C. A. Higgins, and J. M. Howell, "Personal computing: Toward a conceptual model of utilization," MIS Q., vol. 15, no. 1, pp. 125–143, Mar. 1991, doi: 10.2307/249443.

[44] S. Taylor and P. A. Todd, "Understanding information technology usage: A test of competing models," Inf. Syst. Res., vol. 6, no. 2, pp. 144–176, Jun. 1995, doi: 10.1287/isre.6.2.144.

[45] A. Bandura, "Social foundations of thought and action," in The health psychology reader, D. Marks, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[46] E. M. Rogers, Diffusion of innovations. New York: ACM The Free Press, 1995.

[47] I. Alvi, "College students' reception of social networking tools for learning in India: an extended UTAUT model," Smart Learn. Environ., vol. 8, no. 1, pp. 1–18, Dec. 2021, doi: 10.1186/s40561-021-00164-9.

[48] S. A. Salloum, W. Maqableh, C. Mhamdi, B. Al Kurdi, and K. Shaalan, "Studying the social media adoption by university students in the United Arab Emirates," Int. J. Inf. Technol. Lang. Stud., vol. 2, no. 3, pp. 83–95, 2018.

[49] D. Sidik and F. Syafar, "Exploring the factors influencing student's intention to use mobile learning in Indonesia higher education," Educ. Inf. Technol., vol. 25, no. 6, pp. 4781–4796, Nov. 2020, doi: 10.1007/s10639-019-10018-0.

[50] Y. Seedat, S. Roodt, and S. D. Mwapwele, "How South African University Information Systems students are using social media," in Proceedings of The International Conference on Social Implications of Computers in Developing Countries, May 2019, pp. 378–389.

[51] Ministry of Higher Education Malaysia, "Public universities," 2020. https://www.mohe.gov.my/en/download/statistics/2020/493-statistik-pendidikan-tinggi-2020-04-bab-2-universiti-awam/file (accessed Dec. 22, 2021).

[52] J. F. Hair Jr., G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray, "An introduction to structural equation modeling," in Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook, Cham: Springer Nature, 2021, pp. 1–30.

[53] N. F. Richter, R. R. Sinkovics, C. M. Ringle, and C. Schlägel, "A critical look at the use of SEM in international business research," Int. Mark. Rev., vol. 33, no. 3, pp. 376–404, May 2016, doi: 10.1108/IMR-04-2014-0148.

[54] C. Ringle, D. Da Silva, and D. Bido, "Structural equation modeling with the SmartPLS," Brazilian J. Mark., vol. 13, no. 2, pp. 56–73, Oct. 2015, doi: 10.5585/remark.v13i2.2717.

[55] J. F. Hair Jr., G. T. M. Hult, C. M. Ringle, and M. Sarstedt, A primer on partial Least squares structural equation modeling (PLS-SEM), 3rd ed. New York: Sage Publications, 2021.

[56] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," J. Acad. Mark. Sci., vol. 43, no. 1, pp. 115–135, Jan. 2015, doi: 10.1007/s11747-014-0403-8.

[57] C. N. Osakwe, B. Ruiz, H. Amegbe, N. B. Chinje, J. H. Cheah, and T. Ramayah. "A multi-country study of bank reputation among customers in Africa: Key antecedents and consequences," J. Retl. Cons. Serv., vol. 56, no. 5, Sep. 2020, doi: 10.1016/j.jretconser.2020.102182

[58] M. Sarstedt, J. Henseler, and C. M. Ringle, "Multigroup analysis in partial least squares (PLS) path modeling: Alternative methods and empirical results," Adv. Int. Mark., vol. 22, pp. 195–218, Aug. 2011, doi: 10.1108/S1474-7979(2011)0000022012.

[59] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," J. Mark. Res., vol. 18, no. 1, pp. 39–50, Feb. 1981, doi: 10.2307/3151312.

[60] W. Ali, I. I. Alasan, M. H. Khan, S. Ali, J. H. Cheah, and T. Ramayah. "Competitive strategies-performance nexus and the mediating role of enterprise risk management practices: a multi-group analysis for fully fledged Islamic banks and conventional banks with Islamic window in Pakistan," Int J Islamic Middle., vol. 15, no. 1, pp. 125-145, Aug. 2021, doi: 10.1108/IMEFM-06-2020-0310.

[61] J. Cohen, Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.

[62] M. H. Purnomo, “Managing YouTube channel as a virtual da’wah movement for Islamic moderatism,” MUHARRIK J. Dakwah dan Sos., vol. 4, no. 1, pp. 97–109, Jun. 2021, doi: 10.37680/muharrik.v4i01.772.

[63] R. Briandana, C. M. Doktoralina, S. A. Hassan, and W. N. Wan Hasan, “Da’wah communication and social media: The interpretation of millennials in Southeast Asia,” Int. J. Econ. Bus. Adm., vol. 8, no. 1, pp. 216–226, 2020, doi: 10.35808/Ijeba/54.

[64] S. D. Hirsbrunner, “Negotiating the data deluge on YouTube: Practices of knowledge appropriation and articulated ambiguity around visual scenarios of sea-level rise futures,” Front. Commun., vol. 6, p. 11, Feb. 2021, doi: 10.3389/fcomm.2021.613167.

APPENDIX A

Constructs	**Items**	
Performance Expectancy	PE1	I would find YouTube useful for me to acquire Islamic information.
	PE2	Using YouTube increases my daily religious practice.
	PE3	Using YouTube enables me to understand Islamic knowledge more quickly.
	PE4	Using YouTube increases my religious faith.
	PE5	Using YouTube enables me to solve uncertainty in Islamic information.
Effort Expectancy	EE1	Learning to operate YouTube for Islamic information acquisition is easy for me.
	EE2	My interaction with YouTube on Islamic information acquisition would be clear and understandable.
	EE3	I would find YouTube easy to use for religious purposes.
	EE4	It would be easy for me to become skillful at using YouTube to find Islamic information.
	EE5	I would find YouTube easy to use for Islamic information acquisition.
Social Influence	SI1	People who are important to me think I should use YouTube for Islamic information acquisition.
	SI2	People who influence my behavior think I should use YouTube for Islamic information acquisition.
	SI3	My peers have been helpful in using YouTube for Islamic information acquisition.
	SI4	In general, most people have supported using YouTube for Islamic information acquisition.
Facilitating Conditions	FC1	I have the resources necessary to use YouTube.
	FC2	It would be comfortable for me to surf YouTube for Islamic information acquisition.
	FC3	I have the knowledge necessary to use YouTube.
	FC4	YouTube is compatible with any gadgets I use.
	FC5	A specific person is available for assistance with the YouTube difficulties.
Behavioral Intention	BI1	I plan to use YouTube for Islamic information acquisition soon.
	BI2	I predict I would use YouTube every day for religious benefit.
	BI3	I intend to use YouTube more often to acquire Islamic information.
	BI4	I want to share Islamic information on YouTube with my friends.
Actual Usage	Using YouTube for religious information on my own initiatives.	